

Universidad Autónoma de Querétaro

Facultad de Ingeniería

RECONSTRUCCIÓN DE PARÁMETROS COSMOGRÁFICOS CON INTELIGENCIA ARTIFICIAL

Tesis

Que como parte de los requisitos para
obtener el grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta:

Mónica Noelia Castillo Santos

Dirigido por:

Dr. Alberto Hernández Almada

Querétaro, Qro. a 15 de agosto del 2025

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

Tesis

RECONSTRUCCIÓN DE PARÁMETROS COSMOGRÁFICOS CON INTELIGENCIA ARTIFICIAL

Que como parte de los requisitos para obtener el grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta:

Mónica Noelia Castillo Santos

Dirigido por:

Dr. Alberto Hernández Almada

Dr. Alberto Hernández Almada
Presidente

Dr. Josué de Jesús Trejo Alonso
Secretario

Dr. Andras Takacs
Vocal

Dr. Marco Antonio Aceves Fernández
Suplente

Dr. Octavio Cornejo Pérez
Suplente

Centro Universitario Querétaro, Qro.
Fecha de aprobación por el Consejo Universitario
México

A mis padres.

Agradecimientos

A mis padres Mónica y Noé, gracias por hacer de mi formación una prioridad, incluso cuando implicó esfuerzos que muchas veces no se ven pero siempre se sienten.

A mi director de tesis, Dr. Alberto Hernández Almada, le agradezco por su guía y su compromiso con este trabajo.

A mis sinodales, por su disposición para formar parte de este proceso. Sus comentarios enriquecieron el contenido y representaron una oportunidad de aprendizaje.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación por el apoyo brindado a través de la beca nacional para estudios de posgrado, que me permitió concluir este proyecto.

A la Universidad Autónoma de Querétaro, por permitirme realizar mis estudios y brindarme un espacio de trabajo.

Resumen

La discrepancia de 5σ entre los valores calculados de la constante de Hubble obtenidos por el telescopio Planck y por el proyecto SH0ES es uno de los problemas más relevantes en la cosmología moderna. En el presente trabajo se realiza la reconstrucción de los parámetros cosmológicos utilizando redes neuronales FeedForward (FFNN) usando 33 datos observacionales independientes de modelos cosmológicos con el fin de generar predicciones del parámetro de Hubble, $H(z)$, el parámetro de desaceleración, $q(z)$, y en particular, sus valores actuales H_0 , q_0 respectivamente, y el valor del corrimiento al rojo de transición z_T .

Con la finalidad de establecer intervalos de confianza de las variables mencionadas, se realizaron 6 mil entrenamientos independientes. Se construyeron 19 modelos FFNN variando el número de neuronas en las capas ocultas, el valor de la tasa de aprendizaje y el porcentaje de datos para el entrenamiento y además de la implementación de técnicas de regularización.

Basados en las funciones del error cuadrático medio y chi-cuadrada, se seleccionan nueve modelos para reportar el promedio ponderado de las cantidades $H_0 = 68.00^{+0.49}_{-0.49}$ km/s/Mpc, el cual se encuentra a 1.3σ del valor obtenido por el telescopio Planck y a 3.6σ de la misión SH0ES, $q_0 = -0.36^{+0.01}_{-0.01}$ y $z_T = 0.71^{+0.02}_{-0.02}$ son consistentes con las predicciones actuales de modelos cosmológicos. Finalmente, estos resultados confirman que el universo se encuentra en una fase de expansión acelerada.

Abstract

The 5σ discrepancy between the values of the Hubble constant calculated by the Planck telescope and the obtained by the SH0ES project is one of the most important problems in modern cosmology. In this work, we perform the reconstruction of cosmographic parameters using Feed-Forward Neural Networks (FFNN) with 33 observational data points that are independent of cosmological models, to generate predictions of the Hubble parameter $H(z)$, the deceleration parameter $q(z)$, and their current values H_0 and q_0 , as well as the transition redshift z_T .

To establish confidence intervals for the mentioned variables, 6,000 independent trainings were made. A total of 19 FFNN models were constructed by varying the number of neurons in the hidden layers, the learning rate, and the percentage of data used for training, in addition to implementing regularization techniques.

Based on the mean squared error and chi-squared functions, nine models were selected to report the weighted average of: $H_0 = 68.00^{+0.49}_{-0.49}$ km/s/Mpc, which lies 1.3σ away from the value obtained by the Planck telescope and 3.6σ from the SH0ES mission. The values $q_0 = -0.36^{+0.01}_{-0.01}$ and $z_T = 0.71^{+0.02}_{-0.02}$ are consistent with current predictions from cosmological models. Finally, these results confirm that the universe is in a phase of accelerated expansion.

Índice general

Agradecimientos	6
Resumen	8
Abstract	9
Índice general	13
1. Introducción	14
1.1. Motivación	14
1.2. Descripción del problema	15
1.3. Justificación	16
1.4. Hipótesis	17
1.5. Objetivos	17
2. Marco teórico	19
2.1. Estado del Arte	19
2.2. Redes Neuronales	20
2.2.1. Capas y neuronas	20
2.2.2. Pesos	22
2.2.3. Función de activación	22
Función sigmoide	23
Función Rectified Linear Unit (RELU)	23
Scaled Exponential Linear Unit (SELU)	24
Función Exponential Linear Unit (ELU)	25
Función exponencial	26

2.2.4.	Hiperparámetros	27
2.2.5.	Métricas de evaluación	27
2.2.6.	División de los datos	28
	Entrenamiento	29
	Validación	29
	Prueba	29
2.3.	<i>Dropout</i>	29
2.4.	Cosmología	31
2.4.1.	Energía oscura	31
2.4.2.	Materia oscura	32
2.4.3.	Ley de Hubble	32
2.4.4.	Constante de Hubble	33
2.4.5.	Tensión de Hubble	33
2.4.6.	Modelos cosmológicos	34
2.4.7.	Parámetros cosmográficos	35
2.4.8.	Parámetro q_0	36
2.4.9.	Corrimiento al rojo de transición	37
2.4.10.	Distancia comóvil	37
2.4.11.	Observaciones al parámetro de Hubble	37
3.	Materiales y Métodos	39
3.1.	Materiales	39
3.2.	Metodología	40
3.2.1.	Diseño del modelo de red neuronal	41
3.2.2.	Preparación de los datos	41
3.2.3.	Entrenamiento y reconstrucción de parámetros . . .	41
3.2.4.	Generación de múltiples predicciones	42
3.2.5.	Optimización del modelo	42
3.2.6.	Evaluación del desempeño	43
3.2.7.	Validación y comparación	43
4.	Resultados	44

4.1. Modelo 1-50-150-1 sin validación	45
4.2. Modelo 1-100-300-1 sin validación	46
4.3. Modelo 1-150-250-1 sin validación	48
4.4. Modelo 1-250-150-1 con validación 20%	49
4.5. Modelo 1-150-250-1 con validación 100%	50
4.6. Modelo 1-150-250-1 Dropout 20%	52
4.7. Modelo 1-150-250-1 Dropout 10% validación 20%	53
4.8. Modelo 1-100-300-1 Dropout 10% validación 20%	54
4.9. Modelo 1-150-250-1 ($lr=3e-3$)	55
4.10. Modelo 1-150-250-1 ($lr=5e-3$)	57
4.11. Modelo 1-150-250-1 ($lr=8e-3$)	58
4.12. Modelo 1-10-20-1 150 épocas	59
4.13. Modelo 1-10-20-1 300 épocas	61
4.14. Modelo 1-10-20-1 500 épocas	62
4.15. Modelo 1-150-50 150 épocas	63
4.16. Modelo 1-250-150-1 150 épocas	64
4.17. Modelo 1-250-150-1 300 épocas	66
4.18. Modelo 1-250-150-1 <i>dropout</i> 20% 300 épocas	67
4.19. Modelo 1-1024-512-1 150 épocas	68
4.20. Tensión de Hubble	69
4.21. Discusión de resultados	71

5. Conclusiones

76

1

Introducción

1.1. Motivación

La constante de Hubble H_0 , mide la tasa de expansión del universo, es uno de los valores más importantes para la cosmología y la astrofísica. En cosmología, este valor define la escala espacial y temporal del modelo estándar de la cosmología Λ CDM, pues determina la edad del universo y las distancias para calcular la estructura a gran escala [1, 2]. En astrofísica, H_0 es muy importante para ajustar la escala de distancias mediante supernovas, esto permite calcular propiedades como masas de planetas, tasas de transformación estelar, o tasas de crecimiento de núcleos activos y cuásares [3]. Además, influye en la caracterización de lentes gravitacionales [4] y en el análisis de fusiones de estrellas de neutrones detectadas mediante ondas gravitacionales [5]. Uno de los problemas más grandes en la cosmología actual, es la diferencia que existe entre los valores obtenidos por la colaboración Planck y la obtenida por el proyecto *Supernovae, H_0 , for the Equation of State, SH0ES*, la cual es cercana a 5σ , lo que hace que la probabilidad de que estas mediciones pertenecieran a la misma distribución es muy baja, a esta discrepancia se le conoce como tensión de Hubble. Además, obtener una reconstrucción del parámetro de desaceleración, $q(z)$, el cual nos dice si el universo se encuentra en una fase de aceleración o desaceleración. Y su valor en $z = 0$, indica su valor en la actualidad e indica la dinámica del universo. Estos parámetros cosmográficos, no pueden obtenerse por mediciones directas, más bien son calculadas a partir de un observable [6].

Es por eso que se necesitan probar nuevos métodos para obtener valores de los parámetros cosmográficos. Una de estas alternativas son las redes neuronales, las cuales hacen predicciones a partir de datos ob-

servacionales que no depende de modelos cosmológicos, lo que hace que estas estimaciones no tengan sesgos generados por modelos cosmológicos.

1.2. Descripción del problema

Observaciones actuales a supernovas tipo Ia, oscilaciones acústicas de bariones, fondo cósmico de microondas y observaciones directas al parámetro de Hubble, indican en conjunto que el universo se encuentra en una fase de expansión acelerada. Para medir la tasa de expansión del universo, se usan dos técnicas, la primera, calculada por la Agencia Espacial Europea mediante el telescopio Planck, la cual realiza mediciones al fondo cósmico de microondas, obteniendo un valor de 67.4 km/s/Mpc , por otro lado, tenemos la medición que realiza el equipo *SH0ES*; (*Supernovae H_0 for the Equation of State of dark energy*), que realiza mediciones de H_0 haciendo observaciones a supernovas tipo Ia, y da un valor de 73.04 km/s/Mpc . A la discrepancia en estos valores se le conoce como tensión de Hubble. Esta diferencia entre las mediciones no puede considerarse un error experimental. Cuando se comparan 2 mediciones independientes con sus incertidumbres, se le conoce como significancia estadística y para este caso, se tiene un valor cercano a 5σ . Esto indica que la diferencia entre ambos resultados es mucho mayor de lo que podría explicarse por errores estadísticos o sistemáticos en cada método. Una discrepancia de 5σ implica que la probabilidad de que ambas mediciones pertenecieran a la misma distribución sería de 1 entre 1.75 millones. Esta tensión podría deberse a errores sistemáticos aún no identificados, a limitaciones propias del modelo cosmológico estándar o a la presencia de nueva física que modifique la evolución de la expansión cósmica entre el universo temprano y el actual.

Para suavizar la tensión de Hubble es necesario hacer predicciones independientes de modelos cosmológicos, es por eso que se propone el uso de inteligencia artificial, pues estas técnicas permiten realizar predicciones directamente de los datos observacionales, sin imponer modelos teóricos.

1.3. Justificación

La aplicación de inteligencia artificial al estudio de la energía oscura representa un avance importante en la cosmología contemporánea. Para comprender la energía oscura es importante conocer su origen y evolución. Una de las principales áreas de investigación es el uso de algoritmos de inteligencia artificial para predecir los valores del parámetro de Hubble y parámetro de desaceleración, así como los valores actuales de estos y el valor del corrimiento al rojo de transición.

El parámetro de Hubble indica la tasa de expansión del universo, y el parámetro de desaceleración indica si el universo se encuentra en una fase de expansión acelerada. Son dos parámetros fundamentales para comprender la dinámica del universo, ya que con el parámetro de Hubble se miden distancias cósmicas, se determina la edad del universo, se determina la geometría del universo y con el se comparan modelos cosmológicos. Sin embargo, calcularlos a partir de observaciones cosmológicas es complejo debido a la naturaleza de los datos, las distancias cosmológicas, las limitaciones de los instrumentos y la dependencia con modelos cosmológicos. Si bien el valor del parámetro de Hubble se puede medir con datos observacionales, existe una discrepancia en su valor dependiendo de las técnicas usadas en su medición. Por otro lado, el parámetro de desaceleración no se puede medir de manera directa con técnicas actuales, pues requiere un observable para realizar curvas de estimación, es por esto que el uso de inteligencia artificial es de utilidad, pues tienen la capacidad de reconocer patrones para realizar predicciones a partir de imágenes y datos temporales, lo que ha demostrado ser efectivo en tareas de clasificación de eventos en astropartículas, como se muestra en [7], donde se utilizaron redes neuronales convolucionales para mejorar la separación entre señales y fondo en telescopios Cherenkov utilizando información de temporización de los fotones. Además, las redes neuronales convolucionales pueden calcular parámetros cosmológicos con mayor precisión que métodos tradicionales como el espectro de potencia, aún cuando se cuenta con ruido significativo, como se muestra en [8], donde se obtuvo 50% de mejora en la precisión de los parámetros a partir de mapas de lente débil.

El valor de H_0 , el valor de $H(z)$ en $z = 0$, puede obtenerse mediante distintas técnicas observacionales, pero estas no tienen un único valor

resultante: las mediciones del fondo cósmico de microondas y las mediciones a supernovas tipo Ia difieren en aproximadamente 5σ , dando como resultado la tensión de Hubble, uno de los principales problemas en la cosmología actual.

Además, con las predicciones realizadas por las inteligencias artificiales, podemos obtener información sobre los parámetros cosmográficos, sin depender de las restricciones generadas por los modelos cosmológicos. Esto representa un gran avance, pues las investigaciones usualmente dependen de modelos teóricos que se ven restringidos por suposiciones de los mismos [9].

Adicionalmente, en [9], se demuestra que se obtienen resultados comparables con técnicas tradicionales de cosmología en la aplicación de redes neuronales a la reconstrucción de parámetros cosmográficos, sin hacer suposiciones sobre modelos cosmológicos. Es por eso que, el uso de inteligencia artificial además de complementar los métodos tradicionales, proporciona una alternativa para distinguir qué modelo cosmológico se ajusta mejor a los datos observacionales, aportando resultados que pueden complementar o replantear los modelos actuales.

1.4. Hipótesis

En comparación con los métodos cosmológicos tradicionales, el uso de redes neuronales y técnicas de aprendizaje automático reconstruirá los parámetros cosmológicos $H(z)$, $q(z)$ a partir de datos observacionales, sin depender del modelo cosmológico, lo que suavizará la tensión de Hubble.

1.5. Objetivos

Objetivo General

Reconstrucción de los parámetros cosmográficos $H(z)$ y $q(z)$, así como el valor de la constante de Hubble, H_0 , mediante modelos de inteligencia artificial.

Objetivos Particulares

- Diseñar una red neuronal que pueda hacer una reconstrucción de los parámetros cosmográficos $H(z)$ y $q(z)$ a partir de datos observacionales independientes de modelos cosmológicos.
- Establecer cual es la arquitectura que mejor ajusta los datos haciendo variaciones en las neuronas en las capas ocultas, la tasa de apredizaje, cambiando el porcentaje de datos de entrenamiento y agregando técnicas de regularización.
- Determinar mediante las predicciones de la red neuronal los valores actuales de los parámetros cosmográficos H_0 , q_0 y el valor del corrimiento al rojo de transición z_T .
- Evaluar los ajustes de la red neuronales usando la función de pérdida MSE y comparando los resultados con el modelo Λ CDM, para seleccionar los mejores modelos y con ellos calcular el promedio ponderado.

2

Marco teórico

En este capítulo se muestran los fundamentos teóricos necesarios para comprender el presente trabajo. Se explican conceptos básicos de redes neuronales, sus principales componentes y técnicas usadas para evitar sobreentrenamiento. En la segunda parte, se presentan los conceptos más importantes de la cosmología, tales como materia y energía oscura.

2.1. Estado del Arte

En los últimos años, el uso de inteligencia artificial ha cambiado la manera en que se plantean problemas cosmológicos. Esta herramienta permite trabajar con grandes volúmenes de datos, imágenes de baja calidad y ha propuesto nuevas técnicas para calcular los parámetros cosmológicos. Uno de los principales trabajos es [10], en donde se entrenaron redes neuronales convolucionales para clasificar las galaxias de *Sloan Digital Sky Survey* según su morfología, en este trabajo se lograron tiempos mucho menores de clasificación además de alcanzar una precisión del 97 %.

Por otro lado, en [11], se usa un modelo de aprendizaje automático supervisado combinado con inferencia bayesiana para reducir la dimensionalidad de conjuntos de datos cosmológicos a gran escala. Para esto, se hace uso de simulaciones cosmológicas que generan aumento de datos sintéticos, haciendo que el modelo sea más diverso y robusto.

Finalmente, en [9], se realiza la reconstrucción del parámetro de Hubble, $H(z)$, tasa de crecimiento de estructura y módulo de distancia, haciendo uso de redes neuronales y datos observacionales indepen-

dientes de modelos cosmológicos, en los cuales se hizo uso de técnicas de regularización como el *dropout* y obteniendo resultados consistentes con métodos tradicionales.

2.2. Redes Neuronales

Una red neuronal es un modelo computacional basado en el funcionamiento del cerebro. Está compuesto por neuronas que procesan la información que se introduce al modelo, luego la procesan por las diferentes capas ocultas de la red y para entregar una salida.[12]. Las redes neuronales aparecen en 1943, cuando Warren McCulloch y Walter Pitts crearon una neurona artificial basándose en la estructura de las neuronas biológicas [13]. En la década de 1950, Frank Rosenblatt inventó el perceptrón, que es una red neuronal simple con solo una capa que podía reconocer patrones [14]. En 1976, usando backpropagation, se creó un algoritmo que permitía ajustar los pesos de las conexiones de las neuronas en la red [15]. A partir del año 2000, gracias al gran avance de los equipos de cómputo, las redes neuronales obtuvieron mejores resultados y aplicaciones más complejas como procesamiento de lenguaje y visión por computadora, permitiendo la creación del deep learning [16].

A continuación se explican los principales elementos de una red neuronal.

2.2.1. Capas y neuronas

En las redes neuronales, el elemento más simple son las neuronas, a través de ellas fluye la información, procesa la información de entrada, identifica patrones y da una salida. Existen diferentes tipos de neuronas dependiendo de la capa en la que se encuentren, Figura 2.1. A continuación se muestran las capas principales.

- Capa de entrada: es la encargada de recibir los datos iniciales que se procesarán en la red. Cada neurona representa una característica del conjunto de datos y su cantidad depende de la dimensionalidad de la misma.
- Capas ocultas: Son las capas intermedias que encuentran los patrones de los datos. No existe una regla para determinar el número

de neuronas en esta capa, pero de esta depende el rendimiento de la red y el tiempo que se necesita para el entrenamiento.

- Capa de salida: En esta capa se generan los resultados finales de la red. La cantidad de neuronas en la capa de salida depende de las características de la red y de los datos [17].

En la Figura 2.1, podemos observar las conexiones entre las neuronas, las cuales se denominan pesos, $w^{(1)}, w^{(2)}, w^{(3)}$, (ver sección 2.2.2). Cada color en la figura representa una capa diferente en la red neuronal. Los círculos verdes indican las capas de entrada, mientras que los círculos azules representan las capas ocultas. El círculo rojo representa la capa de salida, las funciones $f(x)$ representan las funciones de activación (ver sección 2.2.3).

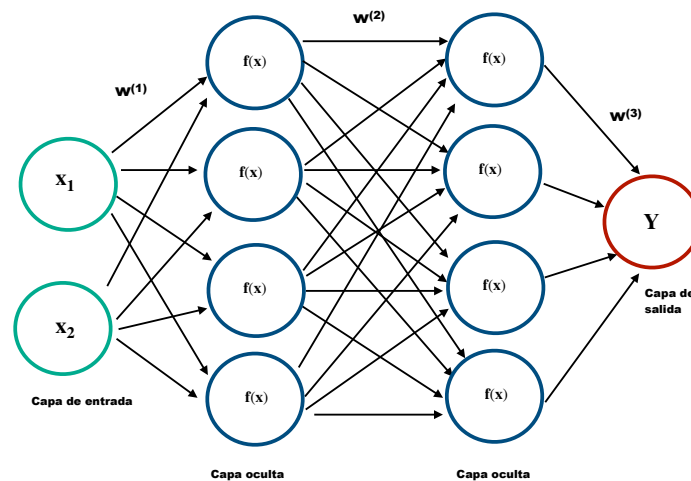


Figura 2.1: Estructura de una red neuronal.

Las redes neuronales se clasifican según su estructura, siendo las principales:

1. Redes neuronales *feedforward*: Conocidas también como redes de propagación hacia adelante, son redes en las cuales la información fluye en solo una dirección, ya que no hay conexión entre neuronas de la misma capa ni con capas anteriores, únicamente con capas siguientes. Este tipo de red neuronal son usadas para clasificación y regresión.

2. Redes neuronales recurrentes: Este tipo de redes neuronales se usan para trabajar con datos secuenciales, es decir, que están ordenados por periodos de tiempo, la principal diferencia con una red neuronal *feedforward* es que tienen la capacidad de recordar información de pasos anteriores, ya que al procesar una entrada, parte de esa información se guarda y se reutiliza en los siguientes pasos, teniendo conexiones hacia atrás y hacia adelante. Son ideales para procesar texto y para predicción de series temporales.
3. Redes neuronales convolucionales: Estas redes neuronales son usadas para procesar datos de más de una dimensión, como imágenes o señales de audio, estas redes fraccionan la información y utilizan un filtro que analiza la imagen por partes [18].

2.2.2. Pesos

Los pesos son parámetros numéricos asociados a cada conexión entre neuronas en las distintas capas, estos son de suma importancia en el aprendizaje, ya que definen cómo las entradas se combinan y se clasifican para calcular la salida de cada una de las neuronas. Conforme la red se entrena, los pesos se ajustan utilizando algoritmos de optimización, como lo es el descenso del gradiente [19].

2.2.3. Función de activación

La función de activación es la unidad básica de procesamiento de una neurona, pues determina si las neuronas serán o no activadas. Si se recibe un dato de entrada en una neurona, se crea una combinación lineal que se multiplica por el peso para después evaluarla en la función de activación y determinar si la neurona se activa y transmite una señal a la neurona de la siguiente capa. La función de activación más sencilla es la lineal, pero esta impide que el modelo prediga patrones de datos complejos[20]. A continuación se describen algunas de las funciones de activación más usadas.

Función sigmoide

Toma una entrada y la convierte en valores entre 0 y 1. Y se define como

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

donde x son los datos.

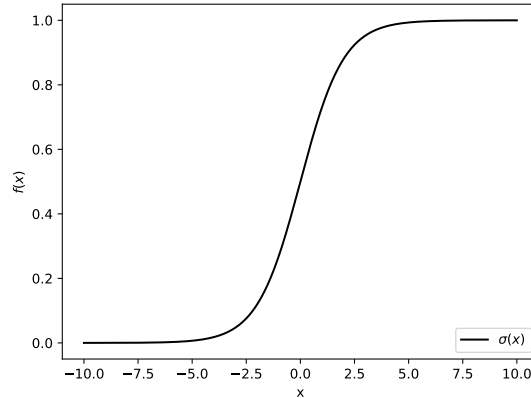


Figura 2.2: Gráfica sigmoide.

La función sigmoide, Figura 2.2, tiene algunas limitaciones que impiden trabajar con cualquier tipo de datos. Entre estos, se encuentra el problema con el descenso del gradiente, pues con valores muy grandes su derivada tiende a cero y así los pesos se actualizan en cantidades muy pequeñas y la red no se entrena de manera correcta.

Función Rectified Linear Unit (RELU)

Esta función tiene una gran eficiencia computacional y soluciona problemas con el descenso del gradiente que tienen otras funciones de activación. La función RELU se define como sigue

$$f(x) = \max(0, x) \quad (2.2)$$

donde \max se define como una función que toma dos valores y devuelve el mayor

$$\max(a, b) = \begin{cases} a, & \text{si } a > b \\ b, & \text{si } b > a \end{cases} \quad (2.3)$$

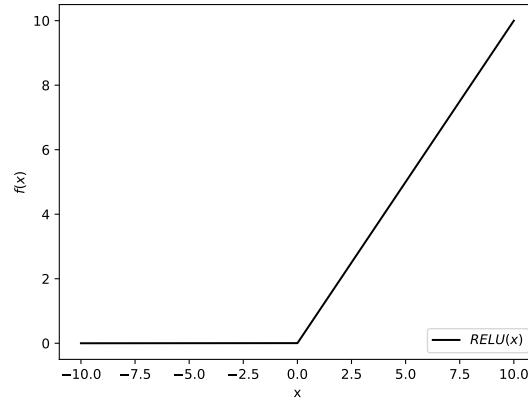


Figura 2.3: Función RELU.

Aunque la función RELU, Figura 2.3, no tiene problemas con datos de gran tamaño y tiende a tener mejor activación en las neuronas con entradas positivas, también cuenta con algunas limitaciones, entre ellas se encuentra el problema de las neuronas muertas, que consiste en que si durante el entrenamiento se obtiene un valor negativo, el gradiente será cero y no activará a neuronas en capas siguientes [21].

Scaled Exponential Linear Unit (SELU)

Esta función de activación, Figura 2.4, nace como una solución a los problemas presentados por la función RELU, pues es una función auto-correctible. La función matemática de SELU es

$$f(x) = \begin{cases} \lambda \cdot (\alpha \cdot e^x - \alpha), & \text{si } x < 0 \\ \lambda \cdot x, & \text{si } x \geq 0 \end{cases} \quad (2.4)$$

donde λ y α son constantes, λ controla la escala de la función y α influye en la forma de la función.

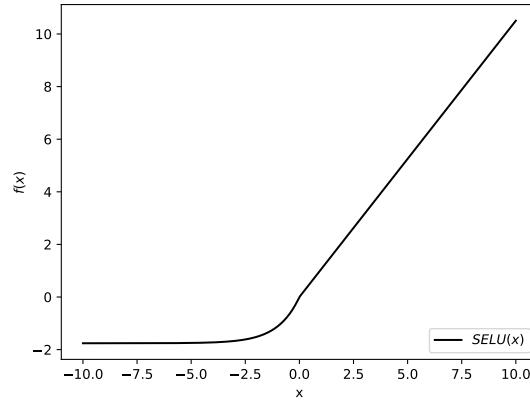


Figura 2.4: Función SELU.

Aunque esta función de activación resuelve los problemas de RELU, no siempre es mejor, ya que depende de los datos de entrada a la red [22].

Función Exponential Linear Unit (ELU)

La función de activación ELU, Figura 2.5, es una variación de la función RELU, con algunas mejoras, siendo diferenciable en todo su dominio y así resolviendo el problema de las neuronas muertas. La ecuación que representa esta función es

$$f(x) = \begin{cases} x, & \text{si } x > 0 \\ \alpha \cdot (e^x - 1), & \text{si } x \leq 0 \end{cases} \quad (2.5)$$

donde α es una constante que determina la pendiente de la función.

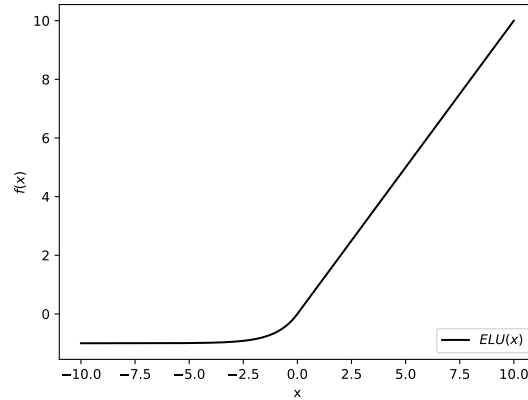


Figura 2.5: Función ELU.

Aunque presenta grandes mejoras, esta función es más compleja de implementar y, por lo tanto, representa más tiempo de cómputo, además de que en muchas ocasiones se presenta un sobreentrenamiento [23].

Función exponencial

Esta función no es tan usada como las mencionadas anteriormente, ya que sufre de problemas con el descenso del gradiente. La ecuación que modela la función exponencial es

$$f(x) = e^x \quad (2.6)$$

En la figura 2.6 se muestra la gráfica para la función exponencial.

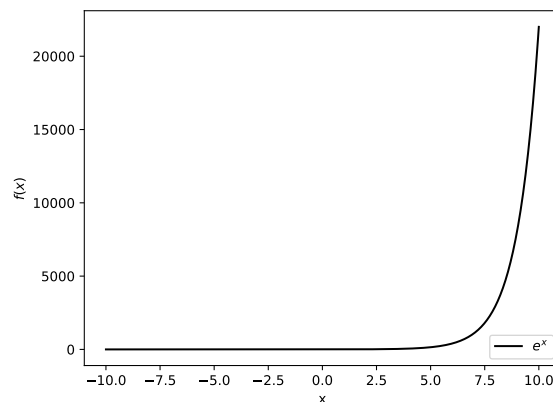


Figura 2.6: Función exponencial

2.2.4. Hiperparámetros

Los hiperparámetros son valores que se determinan antes del entrenamiento de una red neuronal y controlan el aprendizaje, estos parámetros no cambian durante el entrenamiento. Los más comunes son [24] :

1. Tasa de aprendizaje: Determina la actualización de los pesos en cada época del entrenamiento. Es importante encontrar un valor correcto para la tasa de aprendizaje, pues valores muy grandes provocan que el modelo no converja al valor mínimo de la función de pérdida, y para valores muy pequeños el modelo se estanca.
2. Número de épocas: Es el número de iteraciones del entrenamiento. Si se tienen pocas épocas puede derivar en subentrenamiento, y si se tienen demasiadas épocas el modelo puede presentar sobreentrenamiento.
3. Optimizador: Es el encargado de actualizar los pesos de una red neuronal. Los más comunes son el descenso del gradiente (SGD) y Adaptive Moment (ADAM).
4. Función de pérdida: También conocidas como métricas de evaluación, nos dicen que tan bien se ha entrenado el modelo.

2.2.5. Métricas de evaluación

Estas métricas nos indican qué tan precisa es una red neuronal.

1. *Prueba χ^2* : Es un método estadístico que compara la distribución real de los datos con la distribución esperada. Es utilizada principalmente para realizar pruebas de hipótesis y calcular intervalos de confianza. Se calcula como sigue:

$$\chi^2 = \sum_{i=1}^n \frac{[y(z_i) - \bar{y}_i]^2}{\sigma_i^2} \quad (2.7)$$

donde $y(z_i)$ es el modelo evaluado en cada z_i , \bar{y}_i es el valor observado y σ_i el error.

2. *Error absoluto medio (MAE)*: Mide el error de la predicción de un modelo

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.8)$$

donde n es el número total de datos, y_i es el valor real observado para la muestra i y \hat{y}_i es el valor predicho por el modelo.

3. *Error cuadrático medio (MSE)*: Esta enfocado en errores mayores debido al término exponencial. Mide el error cuadrado de la predicción de un modelo.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.9)$$

donde n es el número total de datos, y_i es el valor real observado para la muestra i y \hat{y}_i es el valor predicho por el modelo.

4. *Error relativo cuadrado (RSE)*: Se usa como medida de como varía el error cuadrático medio con respecto a la varianza.

$$RSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.10)$$

donde n es el número total de datos, y_i es el valor real observado para la muestra i y \hat{y}_i es el valor predicho por el modelo.

5. *Coeficiente R^2* : Nos da información sobre la variación de los datos con respecto a los predichos por el modelo. La ecuación para esta métrica es

$$R^2 = 1 - RSE \quad (2.11)$$

Toma valores entre 0 y 1, siendo 1 el modelo ajustado a la perfección a los datos y 0 la nula coincidencia entre los datos y la salida de la red. Para mayor información de las métricas de evaluación, consultar[\[25\]](#).

2.2.6. División de los datos

Para el entrenamiento de una red neuronal, se requiere la división de los datos disponibles, ya que si no se hace esto, el modelo puede sobreajustar los datos de entrenamiento en lugar de aprender patrones generales. Para evitar esto, es necesario dividir los datos en conjuntos

para poder medir el rendimiento del modelo. Los datos se dividen en tres conjuntos: entrenamiento, validación y prueba. A continuación se darán una breve descripción de cada uno [26].

Entrenamiento

Es el conjunto de datos que se utiliza para ajustar los parámetros del modelo y minimizar la función de pérdida. A medida que el conjunto de entrenamiento es más grande, el modelo tiene datos de los cuales aprender patrones importantes, aunque tener más datos no siempre garantiza un mejor desempeño. Generalmente, el tamaño de este conjunto es entre 60 % y 80 % del total de la muestra de datos.

Validación

Este conjunto de datos tiene como objetivo ajustar los hiperparámetros de la red. El modelo no utiliza estos datos para el entrenamiento, sino que los utiliza para evaluar su rendimiento después de cada iteración y ayuda a detectar sobreajuste o subajuste.

Al evaluar el desempeño de datos no usados para el entrenamiento, podemos detectar si el modelo está aprendiendo de manera efectiva o si solo se está adaptando a los datos de entrenamiento. Además, si el modelo funciona bien en los datos de entrenamiento pero no en los datos de validación, puede indicar que el modelo es demasiado complejo para los datos y se debe ajustar la estructura de la red.

El tamaño para este conjunto de datos es entre el 10 % y 20 % del total de los datos [27].

Prueba

Este conjunto se usa para evaluar el rendimiento final del modelo. No se utiliza en el entrenamiento ni en la validación. El tamaño de esta muestra es de entre el 10 % y 20 % del tamaño total de los datos.

2.3. Dropout

El *Dropout* es una técnica de regularización utilizada en el entrenamiento de redes neuronales para evitar el sobreajuste. Funciona desacti-

vando de forma aleatoria un porcentaje de neuronas durante el entrenamiento, haciendo que esas neuronas no participen en el entrenamiento. De esta forma, la red no se vuelve dependiente de ciertas conexiones específicas y aprende a ser más general. Usualmente, el porcentaje de desactivación está entre el 20% y el 30%. Durante la fase de evaluación o prueba, todas las neuronas se mantienen activas, pero sus salidas se ajustan para tener en cuenta que durante el entrenamiento algunas estuvieron apagadas [28].

2.4. Cosmología

La cosmología es la rama de la física que estudia el origen, evolución, composición y posible desenlace del universo a gran escala. Observaciones actuales de Supernovas tipo Ia, oscilaciones acústicas de bariones, fondo cósmico de microondas y observaciones al parámetro de Hubble, indican que el universo se encuentra en una fase de expansión acelerada, este fenómeno es causado por la energía oscura, que representa el 68 % del universo y cuya naturaleza sigue siendo desconocida [29]. Además, se estima que el 27 % del universo es materia oscura, un tipo de materia que no interactúa con radiación electromagnética, y se conoce de su existencia por sus efectos gravitacionales, como la rotación de las galaxias y la formación de estructura a gran escala. En este capítulo se exploran estos conceptos fundamentales en la cosmología.

2.4.1. Energía oscura

La energía oscura es una forma de energía repulsiva cuya naturaleza es desconocida, ya que no interactúa con la radiación electromagnética, lo que la hace indetectable mediante métodos tradicionales. Sin embargo, se sabe de su existencia por la expansión acelerada del universo. En la década de 1920, Edwin Hubble observó que las galaxias se alejaban unas de otras, indicando que el universo se encontraba en expansión. Años más tarde, en 1917, Albert Einstein postuló la constante cosmológica Λ , en un intento de modelar el universo de forma estática, aunque más tarde sugirió que la constante cosmológica podría estar asociada con una presión negativa que impulsa la expansión del universo [30].

El concepto de energía oscura fue propuesto en 1998, cuando observaciones arrojaron la necesidad de un nuevo componente en el universo. Según estudios realizados con el *Hubble Space Telescope*, se estima que la energía oscura ha estado presente durante al menos 9 mil millones de años y constituye aproximadamente el 70 % del contenido del Universo [31].

2.4.2. Materia oscura

La materia oscura es otro de los principales componentes del universo cuya naturaleza es desconocida. Aunque no interactúa directamente con la materia bariónica ni con la radiación electromagnética, se sabe de su existencia por los efectos gravitacionales que ejerce sobre la materia visible, como las estrellas y el gas dentro de las galaxias, así como sobre la dinámica de los cúmulos galácticos. Se estima que la materia oscura constituye aproximadamente el 25 % del universo [32]. La materia oscura fue descubierta por Fritz Zwicky en 1930, cuando se le aplicó el teorema del virial a cúmulos de galaxias, lo que predecía la presencia de materia excedente no visible, para explicar las discrepancias observadas en las velocidades de las galaxias dentro de los cúmulos. [33]. En los años 70, la astrónoma Vera Rubin al estudiar las curvas de rotación de las galaxias espirales, observó que las estrellas en los límites de las galaxias giraban a velocidades más altas de lo que predecía la mecánica clásica, lo que indicaba la presencia de una gran cantidad de masa no visible.

Aunque aún no se conoce el tipo de partículas que componen la materia oscura, se cree que debe estar compuesta por materia oscura fría, para permitir la formación de estructuras a gran escala en el universo. Los neutrinos, por su alta velocidad, han sido descartados como candidatos. Actualmente, los principales candidatos son los axiones o los WIMPs, aunque ninguno de ellos ha sido detectado de manera concluyente hasta el momento [34].

2.4.3. Ley de Hubble

La ley de Hubble establece que la velocidad de recesión de una galaxia es proporcional a la distancia a la que se encuentra, es decir, nos dice que la velocidad a la que una galaxia se aleja de otra es directamente proporcional a la distancia que existe entre ellas y está definida por la siguiente ecuación

$$H = \frac{\dot{a}}{a} \quad (2.12)$$

donde $a = a(t)$ es el factor de escala definido como $a(t) = \frac{1}{1+z}$, el cual está en términos del corrimiento al rojo y \dot{a} es su derivada temporal [31].

El corrimiento al rojo, z , es un tipo de efecto Doppler pero aplicado

a la luz. Se sabe que el universo se encuentra en expansión ya que si las galaxias se estuvieran acercando, la longitud de onda de luz que éstas emiten se correría al azul, caso contrario, si las galaxias se alejan, la luz se corre al rojo.

2.4.4. Constante de Hubble

La constante de Hubble (H_0), representa la tasa de expansión del universo en la actualidad, es decir, qué tan rápido se expande el universo hoy en día. Actualmente, existe un debate sobre el valor exacto del valor de H_0 , dependiendo de las técnicas de medición de esta constante [35].

Se tienen dos principales mediciones a la constante de Hubble, la primera obtenida por la Agencia Espacial Europea mediante el telescopio Planck, el cual realiza mediciones al fondo cósmico de microondas, se obtuvo que el valor de $H_0 \approx 67.4$ km/s/Mpc, por otra parte se tiene el valor obtenido por el equipo *SH0ES* (*Supernova H0 for the Equation of State of Dark Energy*) liderado por Adam Riess, que realiza mediciones de H_0 con supernovas tipo *Ia*, se consiguió el valor de $H_0 \approx 73.04$ km/s/Mpc.

2.4.5. Tensión de Hubble

La tensión de Hubble es la discrepancia que se tiene entre los valores del parámetro de Hubble obtenidos por el telescopio Planck, el cual toma mediciones al fondo cósmico de microondas, y los obtenidos por el proyecto SH0ES, el cual hace mediciones a supernovas tipo *Ia*. Haciendo uso del modelo estándar de la cosmología, se obtiene que, haciendo uso de datos de Planck, el valor de la constante de Hubble es $H_0 \approx 67.4$ km/s/Mpc, mientras que usando datos de SH0ES tenemos un valor de $H_0 \approx 73.04$ km/s/Mpc. A la discrepancia antes mencionada se le conoce como tensión de Hubble y se calcula con la siguiente ecuación [36, 37].

$$\sigma = \frac{M_i - SH0ES}{\sqrt{(e_{M_i})^2 + (e_{SH0ES})^2}} \quad (2.13)$$

donde M_i es el valor de H_0 calculado por cada modelo, *SH0ES* es el valor calculado por *SH0ES*, e_{M_i} es el error de cada modelo y e_{SH0ES} es el error reportado por *SH0ES*.

Esta diferencia, cercana al valor de 5σ , se ha convertido en uno de los principales objetos de estudio en la cosmología moderna, ya que podría indicar la necesidad de modificar el modelo cosmológico actual o introducir nueva física, como componentes exóticos de energía oscura [38].

2.4.6. Modelos cosmológicos

Modelo Λ CDM

El modelo Λ CDM es considerado el modelo estándar de la cosmología, describe la evolución del universo a gran escala. Toma en cuenta tres componentes, Λ , una constante cosmológica relacionada con la energía oscura, materia oscura fría, para permitir la formación de estructura y materia bariónica, es decir, materia visible, tal como átomos, estrellas y planetas.

La ecuación que modela la dinámica del universo es la ecuación de Friedmann, y se expresa en términos del corrimiento al rojo z . Para el modelo, la ecuación de Friedmann está dada por:

$$E^2(z) = \frac{H^2(z)}{H_0^2} = \Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_\Lambda \quad (2.14)$$

donde $\frac{H^2(z)}{H_0^2} = \frac{\dot{a}}{a}$, a es el factor de escala y $\dot{a} = \frac{da}{dt}$, H_0 está relacionado con el parámetro de Hubble de la siguiente manera $H_0 = 100h$, donde h es un parámetro libre del modelo, Ω_{r0} , es la densidad de radiación en la época actual y es del orden de 10^{-5} , mientras que el parámetro Ω_{m0} representa la densidad de materia bariónica y oscura y Ω_Λ representa la densidad de materia oscura [39].

El modelo Λ CDM presenta problemas, como lo son [40]

- El problema de la subestructura: El modelo predice la existencia de una mayor cantidad de subestructura que la observada.
- Problema de la densidad central: Λ CDM predice la existencia de mayor materia oscura en el centro de las galaxias, sin embargo, observaciones realizadas a galaxias espirales desmienten esta predicción.
- El problema de satélite perdido: Esta cosmología contempla la existencia de galaxias enanas que orbitan a otras de mayor tamaño, lo

que no concuerda con las observaciones actuales.

Es por esto que se buscan modelos alternos que solucionen estos problemas, tales como modelos de gravedad modificada, modelos de dimensiones extra, modelos de dinámica newtoniana, modelos con parametrizaciones fenomenológicas.

Modelos alternativos

Estos modelos buscan arreglar los problemas de Λ CDM y proponen diferentes teorías sobre la dinámica del universo, algunos de ellos son:

- *Modelos de parametrizaciones fenomenológicas*
Para estos modelos, se requiere modificar la dinámica del universo mediante la ecuación de estado de la energía oscura, ya que según estudios recientes esta debe ser dinámica, es decir, que varíe con el tiempo y con esto resolver el problema de la tensión de Hubble [41, 42].
- *Modelos de dimensiones extra*
Estas cosmologías predicen la existencia de más dimensiones espaciales. Mencionan que el universo es una brana, el cual es un ente tridimensional que se encuentra inmerso en un espacio de mayor dimensión [43].
- *Modelos de gravedad modificada*
Proponen que la gravedad se modifica a escalas cósmicas, lo que explicaría la expansión acelerada del universo sin tomar en cuenta la energía oscura [44].
- *Modelos de fluidos viscosos*
Estas cosmologías afirman que el universo está inmerso en un fluido con presión negativa, que es el causante de la expansión acelerada del universo [45].

2.4.7. Parámetros cosmográficos

Son un conjunto de parámetros que describen la evolución del universo, son expansiones al parámetro de Hubble a través de serie de Taylor. Estos parámetros se basan en las hipótesis cosmológicas de homogeneidad e isotropía, las cuales nos dicen que, a escalas cosmológicas,

el Universo es uniforme y se expande de manera similar en todas las direcciones.

■ **Parámetro de Hubble**

El parámetro de Hubble, mide la tasa de expansión del universo en función del tiempo

$$H(z) = \frac{\dot{a}}{a} \quad (2.15)$$

conecta los modelos cosmológicos con los datos observacionales.

■ **Parámetro de desaceleración**

$$q(z) = \frac{(z+1)}{H(z)} \frac{dH(z)}{dz} - 1 \quad (2.16)$$

representa la aceleración o desaceleración de la expansión del Universo.

Un valor positivo de $q(t)$ indica una desaceleración en la expansión del Universo, mientras que un valor negativo representa una expansión acelerada. Con este parámetro podemos entender como ha cambiado la dinámica del universo a lo largo del tiempo. En épocas tempranas del universo, la expansión era más lenta debido a la influencia gravitacional de la materia, pero hoy en día la expansión se ha acelerando, teniendo un parámetro de desaceleración negativo.

2.4.8. Parámetro q_0

El valor del parámetro de desaceleración evaluado en $z = 0$ recibe el nombre de q_0 . Este valor presenta el valor del parámetro de desaceleración en la actualidad. El valor de q_0 es importante para saber el estado del universo. Cuando este valor es negativo, indica que el universo se encuentra en una fase de expansión acelerada.

Actualmente, el valor se encuentra entre $-1 < q_0 < -0.5$ [46], lo que quiere decir que el universo se encuentra en una fase de expansión acelerada. Es importante mencionar que el valor de q_0 depende del modelo cosmológico y de los datos observacionales utilizados. El modelo estándar nos dice que $q_0 \approx -0.55$ [41].

2.4.9. Corrimiento al rojo de transición

En la historia del Universo, han existido diferentes etapas en las que diferentes componentes fueron dominantes. Después del Big Bang, la radiación fue el componente que existía en mayor cantidad, con la energía distribuida principalmente en forma de fotones y partículas que se movían a grandes velocidades. A medida que el Universo se expandía y enfriaba, la materia comenzó a dominar, permitiendo la formación de estructura.

Observaciones recientes han demostrado que la energía oscura se ha convertido en el componente dominante del Universo. Este fenómeno, en el que la energía oscura se volvió el componente principal en el Universo, se conoce como corrimiento al rojo de transición de transición, z_T [47].

2.4.10. Distancia comóvil

La distancia comóvil, D_C , describe cómo dos objetos en el universo aumentan su distancia debido a la expansión acelerada. Esto puede calcular la distancia entre dos cuerpos sin importar la expansión del universo, utilizando una unidad de longitud que no varía con la expansión del universo y se calcula como sigue:

$$D_C(z) = c \int_0^z \frac{dz'}{H(z')} \quad (2.17)$$

donde z' es una variable de integración y $H(z')$ depende del modelo cosmológico. Esta distancia también se utiliza para medir cómo viaja la luz en el espacio a través de diferentes épocas del universo. La distancia comóvil al estar en términos de z , podemos explicar cómo la expansión del universo afecta la longitud de onda de la luz [48]

2.4.11. Observaciones al parámetro de Hubble

El parámetro de Hubble indica la tasa de expansión del universo, es decir, la velocidad con la que las galaxias y los objetos en el universo se separan unos de otros.

Para obtener el valor del parámetro de Hubble, existen diversas técnicas como medir radiación proveniente del big-bang, o las mediciones

a Supernovas tipo Ia, pero estas son dependientes de modelos cosmológicos y las limitantes que estos contienen, es por eso que se propone una técnica usando cronómetros cósmicos, en la cual se toma una muestra de 130,000 galaxias masivas y en evolución pasiva, las cuales se encuentran en un rango específico de corrimiento al rojo [49].

La ventaja del uso de cronómetros cósmicos sobre los demás métodos es que las mediciones de parámetro de Hubble son independientes no solo de Λ CDM, sino de cualquier modelo cosmológico, lo que resulta en no causar sesgo en la estimación de los parámetros libres de los modelos cosmológicos [49]. En la figura 2.7, podemos observar los datos de [50] graficados con sus respectivas barras de error.

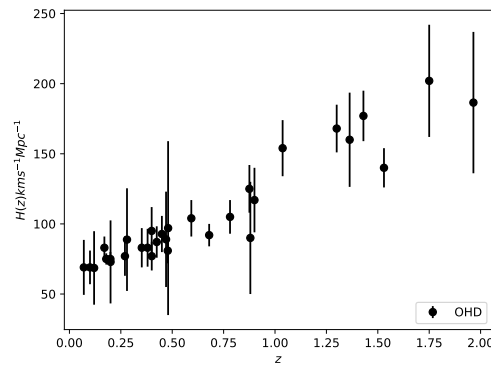


Figura 2.7: Datos observacionales del parámetro de Hubble

3

Materiales y Métodos

En este capítulo se presenta la metodología empleada para la reconstrucción de los parámetros cosmográficos $H(z)$ y $q(z)$ a partir de datos observacionales, utilizando técnicas de inteligencia artificial, ya que se implementa una red neuronal *feedforward* (FFNN) para lograr ajustes con las mediciones disponibles.

Se describen el conjunto de datos utilizados, así como el diseño, la selección y evaluación de distintas arquitecturas de red neuronal. Además, se describen las distintas fases que se siguieron para llegar al resultado.

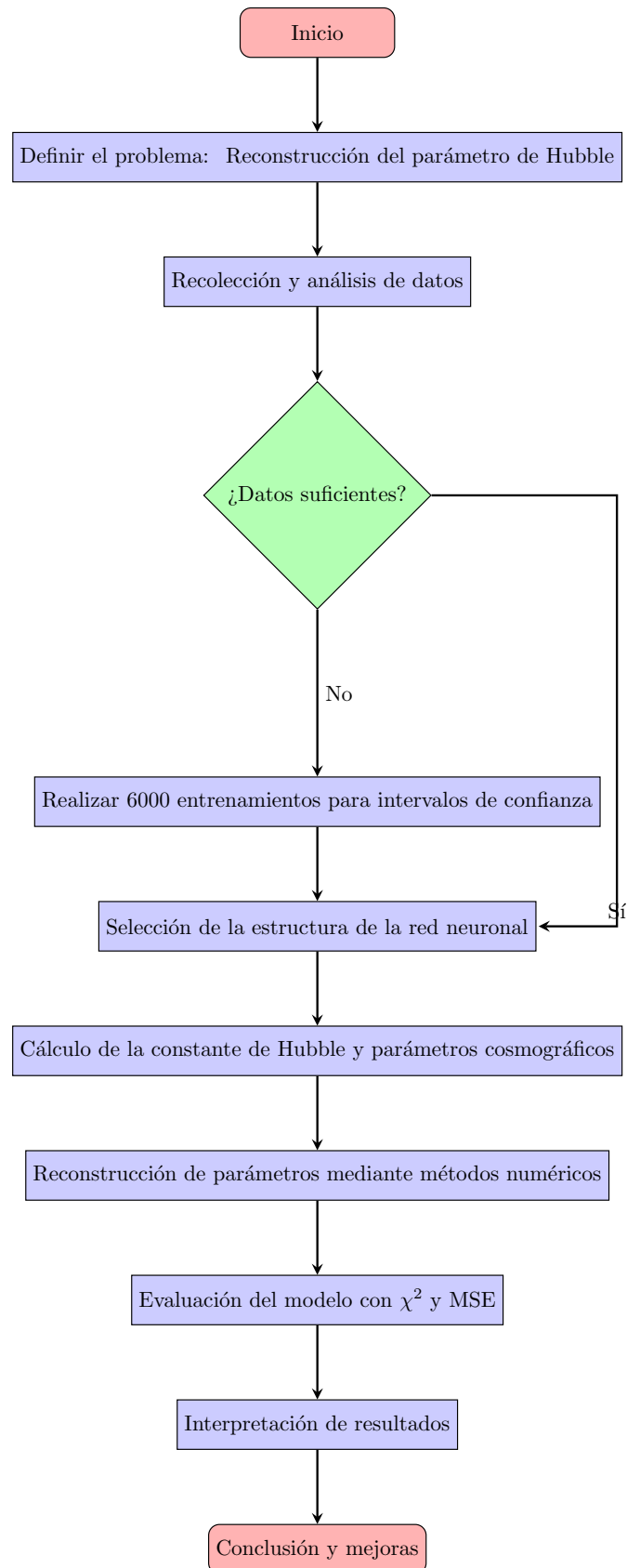
Cada uno de los resultados se evaluó mediante métricas estadísticas como el error cuadrático medio (MSE) y χ^2 . Finalmente, se presenta el diagrama de flujo que resume el enfoque seguido en esta investigación.

3.1. Materiales

Los recursos materiales a usar en este proyecto son:

- Computadora personal: MacBook Pro con chip M3 Pro, CPU de 12 núcleos, GPU de 18 núcleos, 18GB de memoria RAM unificada y SSD de 512GB.
 - Base de datos pública [50] Observaciones al parámetro de Hubble.
-

3.2. Metodología



El objetivo principal de este trabajo es implementar un modelo de red neuronal que permita reconstruir el parámetro de Hubble $H(z)$ y de $q(z)$ a partir de datos observacionales, para así poder suavizar la tensión de Hubble. A continuación se describen las etapas metodológicas llevadas a cabo:

3.2.1. Diseño del modelo de red neuronal

Se probaron diversas arquitecturas con el fin de encontrar la red neuronal que mejor se ajustara a los datos.

- Se probaron arquitecturas con diferentes combinaciones de capas ocultas y número de neuronas.
- Se usó la técnica de regularización *dropout* en ciertos modelos, para evitar el sobreajuste, ya que se contaba con poca cantidad de datos observacionales.
- Se modificaron los hiperparámetros de la red neuronal como la tasa de aprendizaje y las épocas durante las que se entrenaron los modelos.

Se eligió una red neuronal con 2 capas ocultas, con función de activación ELU y sigmoide respectivamente, el optimizador ADAM y una función de pérdida MSE.

3.2.2. Preparación de los datos

Ya que solo se contaba con 33 datos observacionales, se realizaron pruebas con modelos entrenados utilizando el conjunto completo. Sin embargo, para algunos experimentos se realizó una división creando conjuntos de entrenamiento y prueba.

3.2.3. Entrenamiento y reconstrucción de parámetros

Una vez definido el modelo, se entrenó la red neuronal y se obtuvo la reconstrucción de $H(z)$ para después calcular los siguientes parámetros cosmológicos:

- El valor del parámetro de Hubble actual, H_0 .
- El parámetro de desaceleración $q(z)$.
- El redshift de transición z_T .

3.2.4. Generación de múltiples predicciones

Para reducir la incertidumbre dado el limitado número de datos, se entrenó cada modelo 6 mil veces:

- Se realizaron entrenamientos con diferentes condiciones iniciales aleatorias y conjuntos de entrenamiento.
- En total, se realizaron 6000 entrenamientos independientes, lo que permitió construir intervalos de confianza sobre las predicciones de los parámetros cosmográficos.

Basados en el experimento presentado en la Tabla 3.1, se determinó que utilizar 6000 entrenamientos es óptimo, ya que para este tamaño se obtuvo uno de los menores valores de χ^2 y de MSE, con una dispersión de errores reducida y una distribución estable

Tabla 3.1: Resultados para distintos tamaños del conjunto de datos

Número de datos	H_0	χ^2	q_0
1500	$68.26^{+1.89}_{-1.19}$	$17.80^{+3.32}_{-0.84}$	$-0.37^{+0.5}_{-0.3}$
3000	$68.29^{+1.92}_{-1.71}$	$17.52^{+3.05}_{-0.81}$	$-0.37^{+0.5}_{-0.3}$
4500	$68.30^{+1.68}_{-1.61}$	$17.53^{+3.05}_{-0.81}$	$-0.37^{+0.5}_{-0.3}$
6000	$68.31^{+1.54}_{-1.55}$	$17.54^{+3.08}_{-0.78}$	$-0.37^{+0.5}_{-0.3}$
7500	$68.32^{+1.77}_{-1.32}$	$17.51^{+3.11}_{-0.79}$	$-0.37^{+0.5}_{-0.3}$
9000	$68.33^{+1.53}_{-1.45}$	$17.47^{+3.08}_{-0.77}$	$-0.37^{+0.5}_{-0.3}$

3.2.5. Optimización del modelo

Se probaron distintas configuraciones sobre el desempeño del modelo:

- Se analizó cómo el número de capas, neuronas, el uso de *dropout* y variación del valor de la tasa de aprendizaje, para detectar como afectaban el error cuadrático medio (MSE).
- Además se realizaron pruebas con arquitecturas más simples, es decir, menor número de neuronas en cada una de las capas ocultas.

3.2.6. Evaluación del desempeño

Para evaluar cuantitativamente el ajuste, se utilizaron dos métricas principales:

- χ^2 , para medir que tan bueno es el ajuste de las predicciones respecto a los datos observacionales.
- El error cuadrático medio, para medir la precisión y estabilidad de las predicciones del modelo.

3.2.7. Validación y comparación

Finalmente, se evaluaron los resultados obtenidos mediante comparación con modelos cosmológicos estándar:

- Se compararon las predicciones del modelo con los valores obtenidos con Λ CDM.
- Se seleccionaron las mejores arquitecturas de red neuronal basado en los valores obtenidos de χ^2 y MSE.

4

Resultados

En este capítulo se presentan los resultados obtenidos en cada uno de las pruebas usando diferentes arquitecturas de redes neuronales para la reconstrucción de los parámetros cosmológicos. Cada modelo se evaluó tomando en cuenta las métricas MSE y χ^2 y comparando los parámetros cosmológicos con los obtenidos con Λ CDM. El valor de $\chi^2_{\Lambda\text{CDM}} = 15.24$ se toma como referencia del ajuste de Λ CDM, ya que con esta métrica podemos determinar que tan bueno es el ajuste de cada uno de los modelos probados. Para tener una referencia adicional de comparación de las predicciones obtenidas de las pruebas con el modelo estándar, se calculó para el valor de MSE para Λ CDM y se obtuvo el valor $MSE_{\Lambda\text{CDM}} = 140.93$

En la tabla 4.1 se muestran los modelos de red neuronal usados en este trabajo, en donde se incluye el nombre que se le dio a cada modelo, número de capas ocultas, tasa de *dropout*, valor del *learning rate*, porcentaje del conjunto de validación y épocas durante las que se entrenó cada modelo.

Tabla 4.1: Configuración de los modelos

Modelo	Neuronas por capa	Dropout	Tasa de aprendizaje	Validación	Épocas
M1	[50, 150]	No	1×10^{-3}	No	150
M2	[100, 300]	No	1×10^{-3}	No	150
M3	[150, 250]	No	1×10^{-3}	No	150
M4	[250, 150]	No	1×10^{-3}	20%	150
M5	[150, 250]	No	1×10^{-3}	100%	150
M6	[150, 250]	20%	1×10^{-3}	No	150
M7	[150, 250]	10%	1×10^{-3}	20%	150
M8	[100, 300]	10%	1×10^{-3}	20%	150
M9	[150, 250]	No	3×10^{-3}	20%	150
M10	[150, 250]	No	5×10^{-3}	20%	150
M11	[150, 250]	No	8×10^{-3}	20%	150
M12	[10, 20]	No	1×10^{-3}	20%	150
M13	[10, 20]	No	1×10^{-3}	20%	300
M14	[10, 20]	No	1×10^{-3}	20%	500
M15	[150,50]	No	1×10^{-3}	20%	150
M16	[250,150]	No	1×10^{-3}	20%	150
M17	[250,150]	No	1×10^{-3}	20%	300
M18	[250,150]	20%	1×10^{-3}	20%	300
M19	[1024,512]	20%	1×10^{-3}	20%	150

4.1. Modelo 1-50-150-1 sin validación

El modelo M1, con arquitectura 1-50-150-1, mostró un desempeño adecuado al estimar los parámetros cosmológicos a partir de los datos observacionales. Este modelo arrojó un valor de $H_0 = 67.25^{+0.57}_{-1.72}$ km/s/Mpc, el cual se acerca a los resultados de Planck. El valor obtenido de $\chi^2 = 16.35$ es comparable al del modelo Λ CDM. La Figura 4.1 muestra la reconstrucción de los parámetros cosmológicos para esta configuración.

El comportamiento de la función de pérdida durante el entrenamiento, Figura 4.1(b), indica una convergencia estable, aunque aún en el fin del entrenamiento esta continúa oscilando. El histograma de los valores reconstruidos de H_0 , Figura 4.1(a), ofrece información sobre la estabilidad estadística del modelo. La distribución presenta una forma simétrica y centrada en el valor medio. Para la reconstrucción de $H(z)$ y $q(z)$, Figuras 4.1(c) y 4.1(d), donde las bandas de incertidumbre se mantienen estrechas y congruentes con los datos observacionales, tenemos que la reconstrucción de $q(z)$ indica que en el entrenamiento logra pre-

decir la transición del universo de desaceleración a aceleración, con un valor $z_T = 0.67$, el cual coincide con las observaciones actuales.

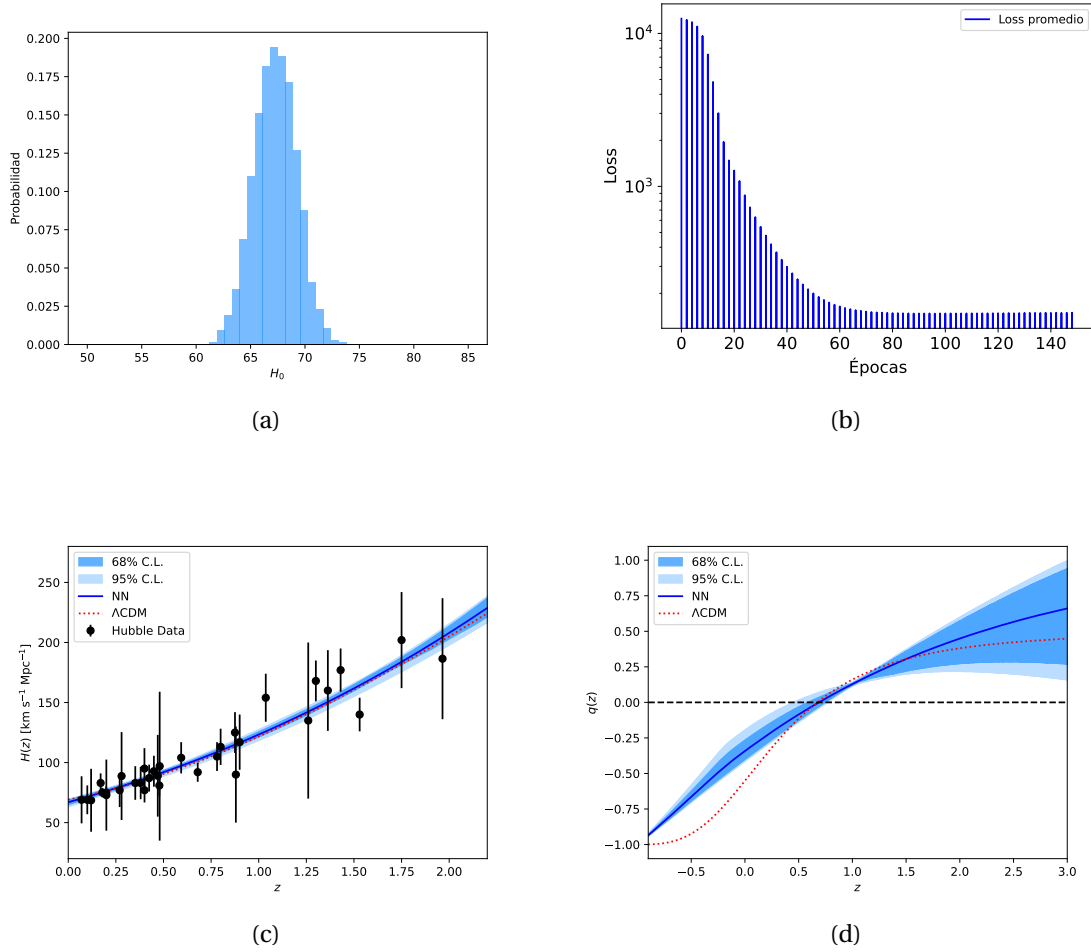


Figura 4.1: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.2. Modelo 1-100-300-1 sin validación

El modelo M2, con arquitectura 1-100-300-1, presentó un buen desempeño tomando en cuenta los valores que se obtuvieron de MSE y χ^2 . Este modelo obtuvo un valor de $H_0 = 68.32_{-1.83}^{+2.27} \text{ km/s/Mpc}$, el cual es consistente con las estimaciones de Planck. El valor de $\chi^2 = 17.80$ es ligeramente mayor al del modelo Λ CDM, pero sigue siendo bueno en comparación con otras configuraciones más complejas. La Figura 4.2 muestra

la reconstrucción de los parámetros cosmológicos para esta configuración.

En la Figura 4.2(b), se observa que la función de pérdida decrece durante las primeras épocas y se estabiliza en valores bajos, aunque muestra mayor dispersión respecto al modelo M1, lo cual puede estar asociado al tamaño y profundidad de la arquitectura. El histograma de H_0 (Figura 4.2(a)) muestra una distribución centrada pero con mayor dispersión. Para las reconstrucciones de $H(z)$ y $q(z)$, Figuras 4.2(c) y 4.2(d), las bandas de error se hacen ligeramente más anchas respecto a M1, aunque siguen la tendencia de Λ CDM. En particular, la evolución de $q(z)$ permite identificar la transición de desaceleración a aceleración, con un valor de $z_T = 0.72$.

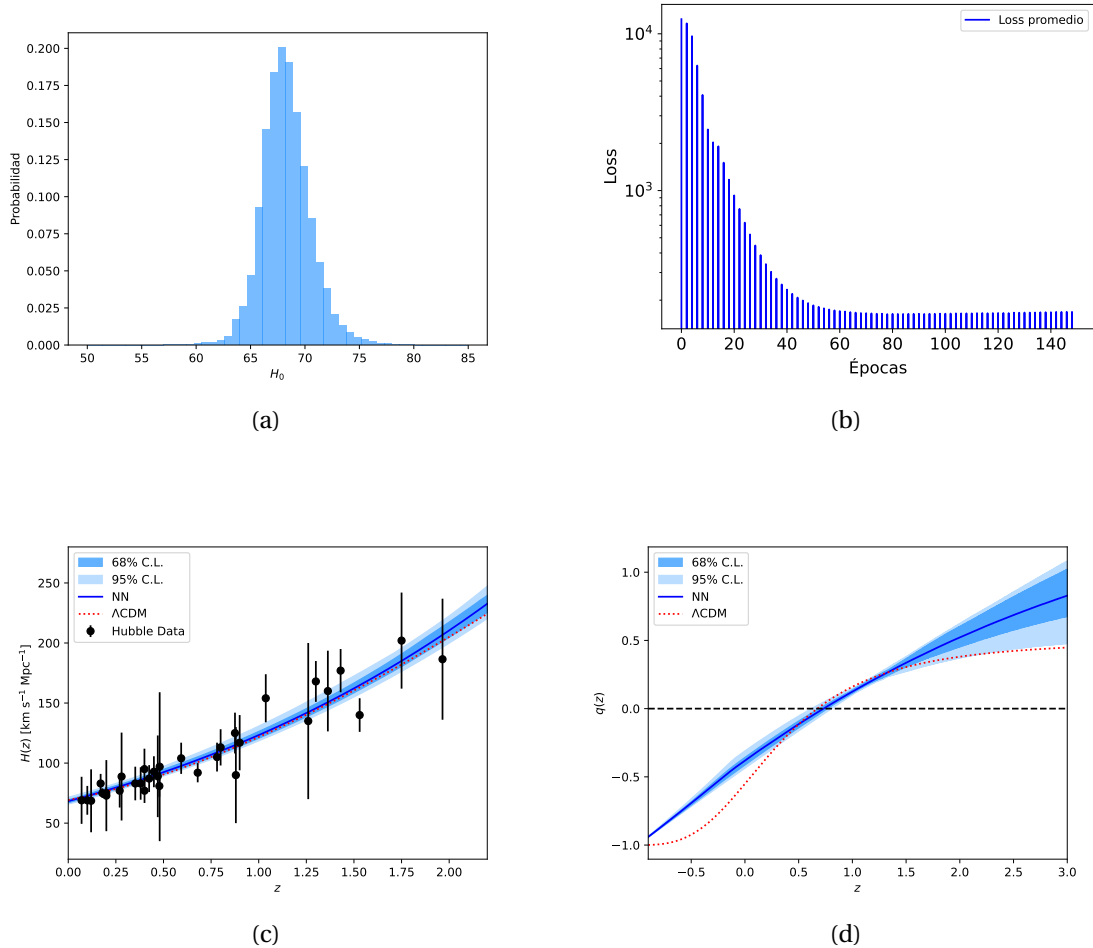


Figura 4.2: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.3. Modelo 1-150-250-1 sin validación

El modelo M3, con arquitectura 1-150-250-1, presentó un comportamiento similar al de modelos anteriores en la reconstrucción de los parámetros cosmológicos. Obtuvo un valor de $H_0 = 68.33^{+1.53}_{-1.45}$ km/s/Mpc, el cual tiende a los valores de Planck. El valor de $\chi^2 = 17.47$ es cercano al del modelo Λ CDM y similar al reportado para M2. La Figura 4.3 presenta la reconstrucción de los parámetros cosmológicos para esta configuración.

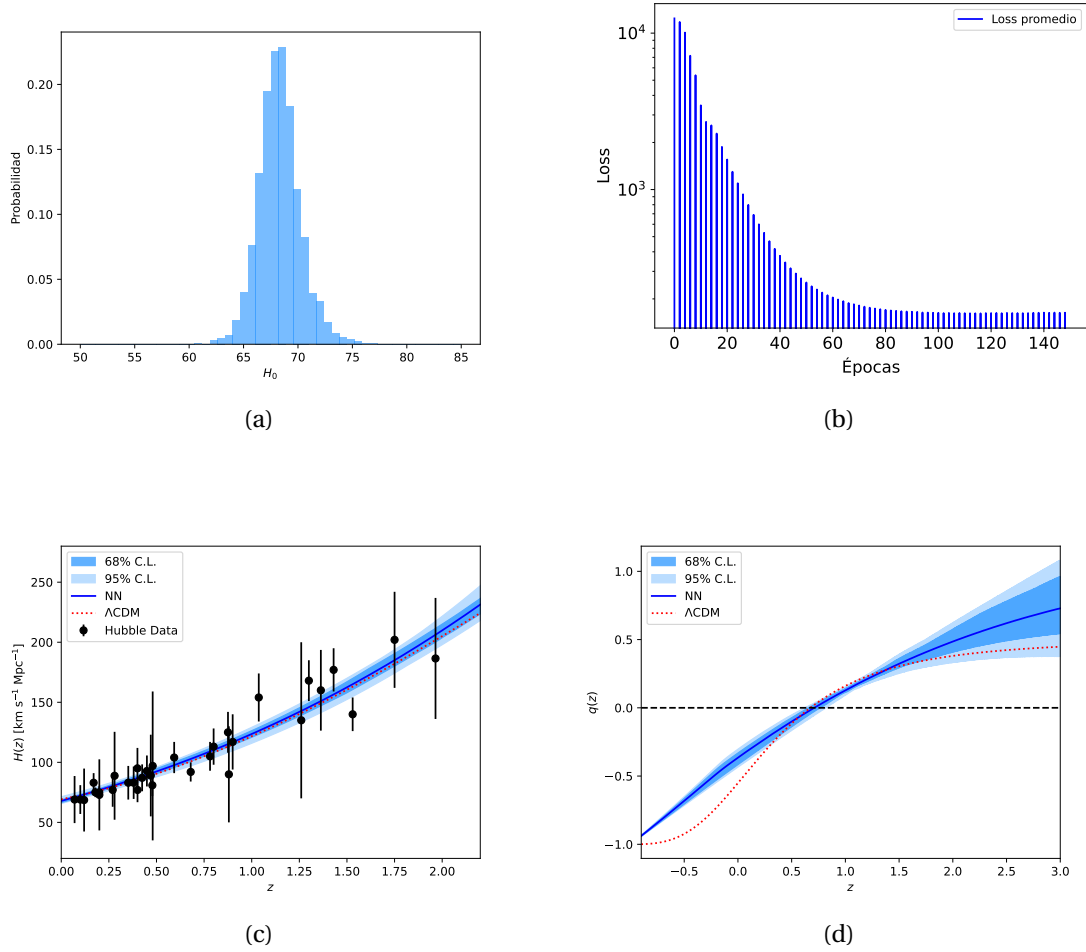


Figura 4.3: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

El comportamiento de la función de pérdida, Figura 4.3(b), muestra una convergencia en todo el entrenamiento, aunque la pérdida tien-

de a estabilizarse de forma más lenta respecto a configuraciones más pequeñas y, al igual que los modelos anteriores, presenta oscilaciones aún en el final del entrenamiento. El histograma de los valores de H_0 , Figura 4.3(a), muestra una distribución simétrica, aunque más estrecha en comparación con modelos como M1. Las reconstrucciones de $H(z)$ y $q(z)$, en las Figuras 4.3(c) y 4.3(d), se ajustan adecuadamente a los datos observacionales. Para este caso $q(z)$ el cual nos da el valor de la transición de la expansión del universo, con un valor de $z_T = 0.72$, concuerda con lo reportado en la literatura.

4.4. Modelo 1-250-150-1 con validación 20 %

Con el modelo M4, con arquitectura 1-50-150-1 y validación del 20 %, se obtuvieron resultados consistentes en la reconstrucción de los parámetros cosmográficos. El valor obtenido de $H_0 = 68.19^{+1.34}_{-1.76}$ km/s/Mpc es cercano a los valores reportados por Planck, mientras que la métrica $\chi^2 = 16.80$, se mantiene cercana al modelo Λ CDM. La Figura 4.4 muestra los resultados de la reconstrucción de los parámetros cosmográficos obtenidos con esta configuración.

En la Figura 4.4(b) se observa que la función de pérdida decrece de forma continua y sin oscilaciones en el entrenamiento alcanzando valores bajos, lo que indica que el modelo converge de forma satisfactoria. El histograma de H_0 , Figura 4.4(a), muestra una distribución más concentrada en comparación con los modelos anteriores, lo cual hace un intervalo de incertidumbre más pequeño. Para los parámetros $H(z)$ y $q(z)$, mostrados en las Figuras 4.4(c) y 4.4(d), las bandas de incertidumbre se redujeron. En el parámetro de desaceleración, reconstruyendo un valor de $z_T = 0.71^{+0.07}_{-0.12}$, podemos observar que la transición de desaceleración a aceleración se hace ligeramente más tarde, aunque estos valores siguen siendo consistentes con los valores cosmológicos actuales.

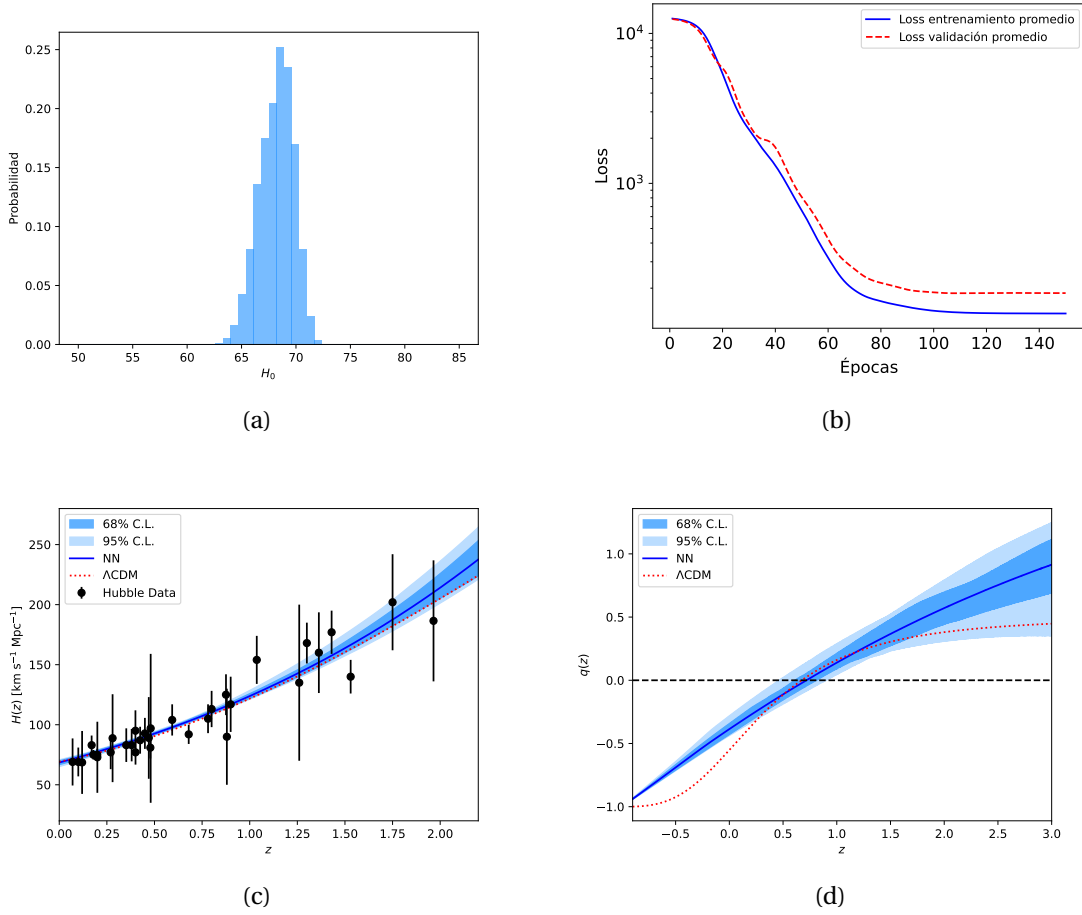


Figura 4.4: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.5. Modelo 1-150-250-1 con validación 100%

El modelo M5, con arquitectura 1-150-250-1 y conjunto de validación igual al conjunto completo de entrenamiento, resultó en una estimación menos precisa de los parámetros cosmológicos en comparación con modelos anteriores. El valor del parámetro de Hubble que se obtuvo fue $H_0 = 68.16^{+1.30}_{-1.97} \text{ km/s/Mpc}$ presenta una dispersión. Además de que las métricas $\chi^2 = 16.16$ y $MSE = 145.82$ muestran que el modelo no llegó a los valores mínimos esperados. La Figura 4.5 muestra los resultados de esta arquitectura.

La función de pérdida, Figura 4.5(b), decrece durante todo el en-

trenamiento, aunque se estabiliza a partir de las 100 épocas sin mejoras considerables y se pueden observar ligeras oscilaciones, lo cual puede indicar sobreajuste y podemos observarlo en la forma del histograma de H_0 , Figura 4.5(a), el cual muestra una distribución más ancha y asimétrica. Las reconstrucciones de $H(z)$ y $q(z)$, en las Figuras 4.5(c) y 4.5(d), presentan bandas de incertidumbre más amplias. Aunque se logra identificar la transición de desaceleración a aceleración, el valor estimado de $z_T = 0.70$ muestra un ligero desplazamiento, pero permanece dentro de un intervalo razonable.

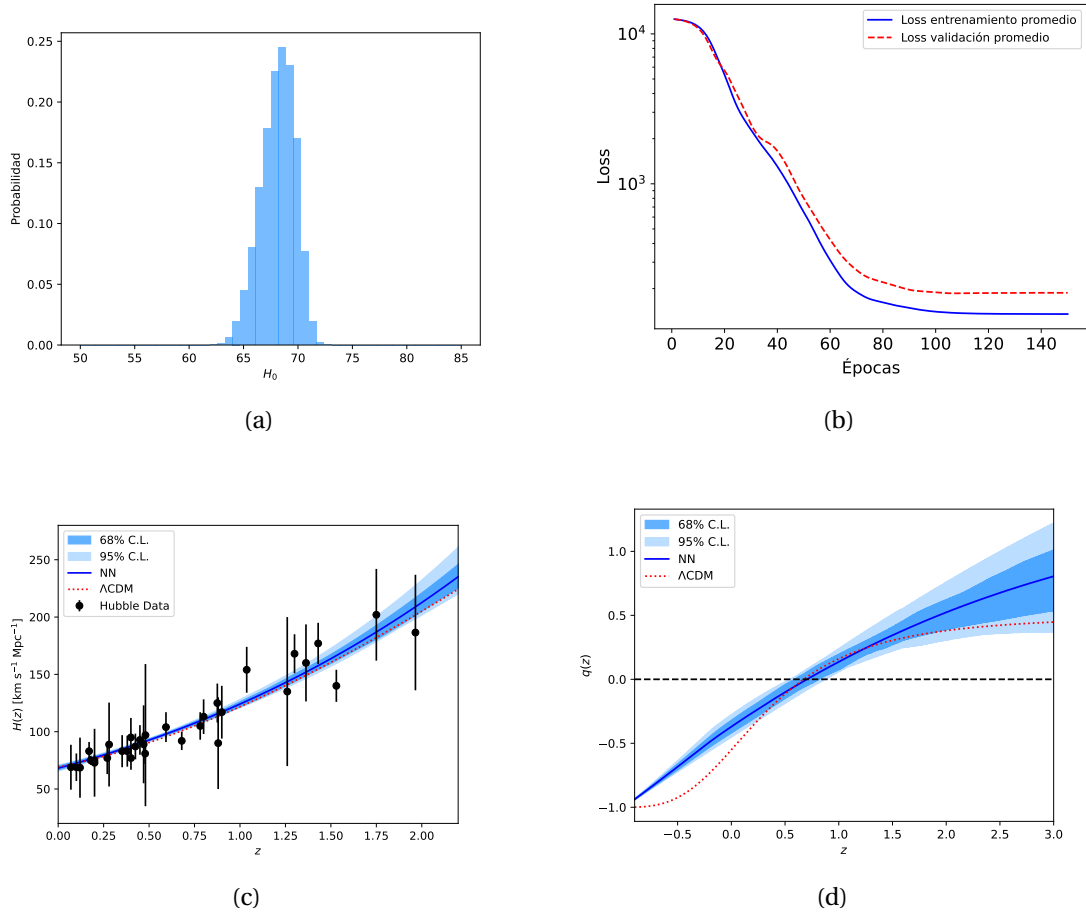


Figura 4.5: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.6. Modelo 1-150-250-1 Dropout 20%

En el modelo M6, se hizo uso de dropout, en donde se apagan de forma aleatoria 20% de las neuronas, y el cual presentó peores resultados en comparación con los modelos anteriores. El valor del parámetro de Hubble reconstruido fue $H_0 = 44.25^{+7.02}_{-6.43}$ km/s/Mpc, el cual no es consistente con lo reportado por las misiones Planck y SH0ES. Además, los valores $\chi^2 = 288.14$ y $MSE = 2096.76$ se encuentran muy elevados, con esto se puede inferir que el modelo no ajustó correctamente a los datos observacionales. En la Figura 4.6 observamos los resultados de esta configuración.

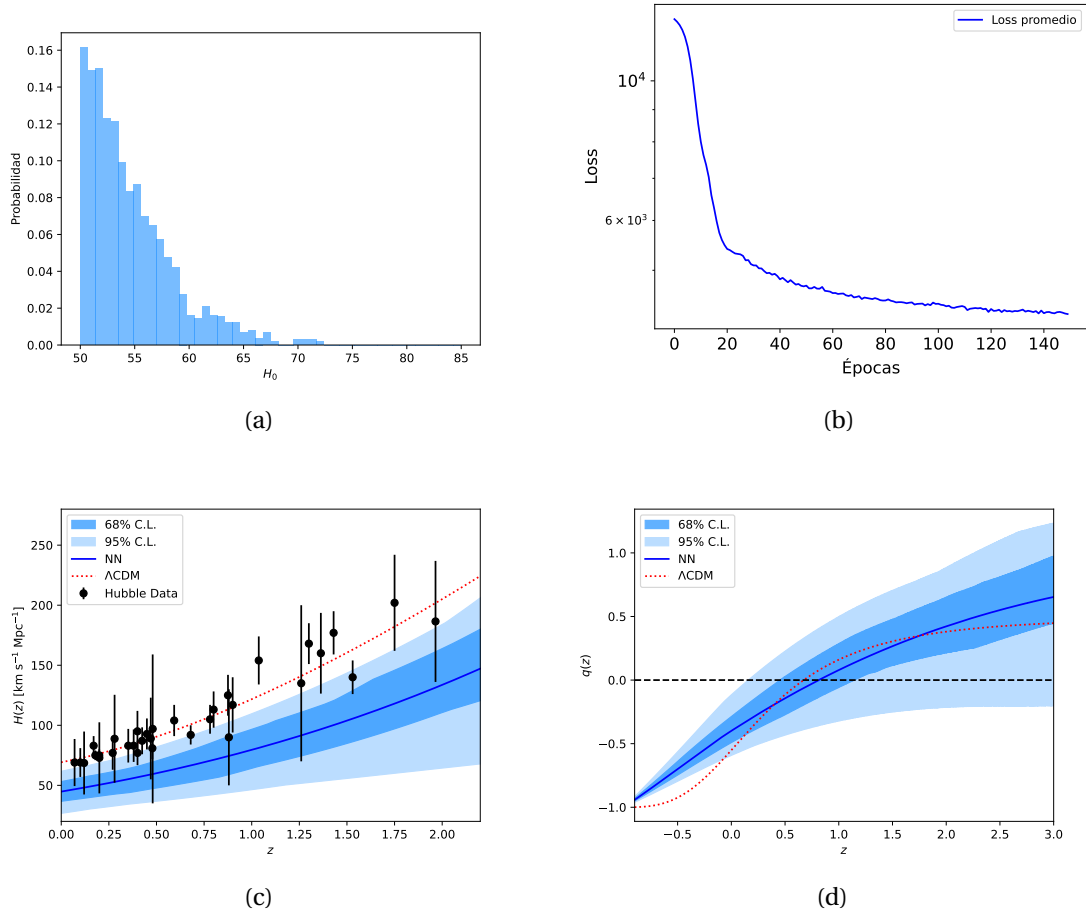


Figura 4.6: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

En la Figura 4.6(b), la función de pérdida decrece en todo el entrenamiento, pero no alcanza valores mínimos, además de que presenta oscilaciones. En el histograma, Figura 4.6(a), se observa que los valores tienden a valores bajos, por lo que muchos de ellos no se encuentran en el rango esperado. Las reconstrucciones de $H(z)$ y $q(z)$, mostradas en las Figuras 4.6(c) y 4.6(d), muestran estas inconsistencias, el modelo se desvía de los datos observacionales y del modelo Λ CDM y con bandas de error muy anchas y por debajo de los valores esperados. Aunque se tiene una transición de desaceleración a aceleración, el valor de $z_T = 0.95$, nos dice que la transición se hizo después de lo que indican estudios actuales.

4.7. Modelo 1-150-250-1 Dropout 10% validación 20%

En el modelo M7, además de incluirse un dropout del 10% se incluyó un conjunto de validación del 20%. Aunque el histograma de los valores predichos para H_0 , 4.7(b), se muestra un poco más en el rango aceptado, se tiene que el valor promedio calculado es $H_0 = 55.80^{+6.66}_{-10.43}$ km/s/Mpc, el cual está aún por debajo de los valores de Planck y SH0ES, además los valores de $\chi^2 = 111.21$ y $MSE = 883.93$, muestran una mejora respecto al modelo M8 en donde no se incluyó conjunto de validación pero sin llegar a los valores mínimos. La Figura 4.7 muestra los resultados de esta configuración.

En la Figura 4.7(b), la función decrece, pero no llega a valores óptimos, además de que presenta oscilaciones. En las reconstrucciones de $H(z)$ y $q(z)$, en las Figuras 4.7(c) y 4.7(d), el valor promedio de la reconstrucción del parámetro de Hubble se encuentra muy por debajo del valor de Λ CDM, adicionalmente, en ambas figuras se observan barras de error anchas, y para el caso de $q(z)$, observamos la transición de aceleración a desaceleración en $z_T = 0.89^{+0.34}_{-0.22}$.

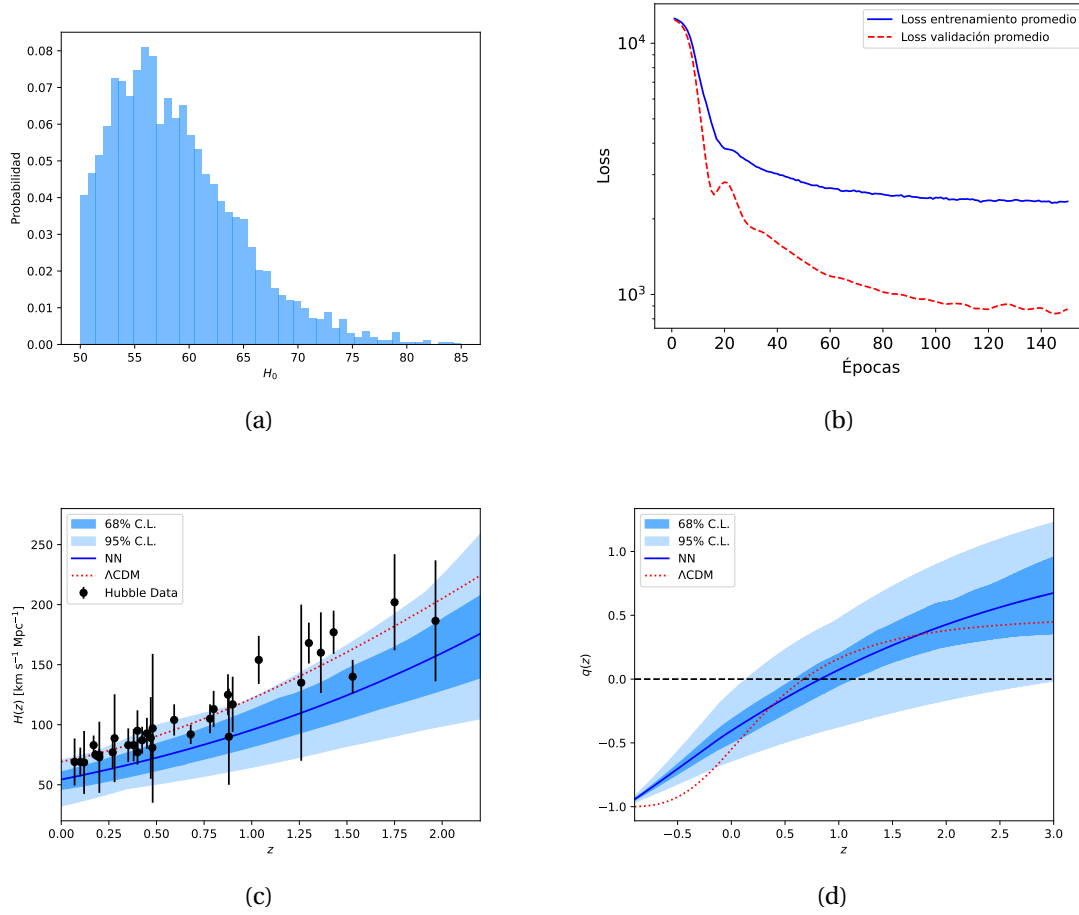


Figura 4.7: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.8. Modelo 1-100-300-1 Dropout 10% validación 20%

En el modelo M8, se tiene la misma configuración que el modelo M7 en cuanto a dropout y conjunto de validación, pero en este caso se tienen más neuronas en las capas ocultas, con la finalidad de ver si con el aumento de las neuronas se tiene un mejor desempeño de la red neuronal. Se obtuvo un valor de $H_0 = 44.84^{+7.76}_{-7.99} \text{ km/s/Mpc}$, el cual se encuentra aún por debajo de los valores de SH0ES y Planck. En los valores de las métricas, $\chi^2 = 294.93$ y $MSE = 2147.24$ no se muestra mejora significativa en comparación al modelo anterior con menos neuronas. La Figura 4.8 muestra los resultados de esta configuración.

La función de pérdida, en la Figura 4.8(b), se muestra un comportamiento similar al de modelos anteriores, aunque en la Figura 4.8(a) se ve la tendencia hacia valores bajos de H_0 . Las reconstrucciones de $H(z)$ y $q(z)$, mostradas en las Figuras 4.8(c) y 4.8(d), tienen bandas de incertidumbre grandes. Se tiene que el valor de transición de $q(z)$ se encuentra en $z_T = 0.99$, este valor es elevado en comparación con cálculos recientes.

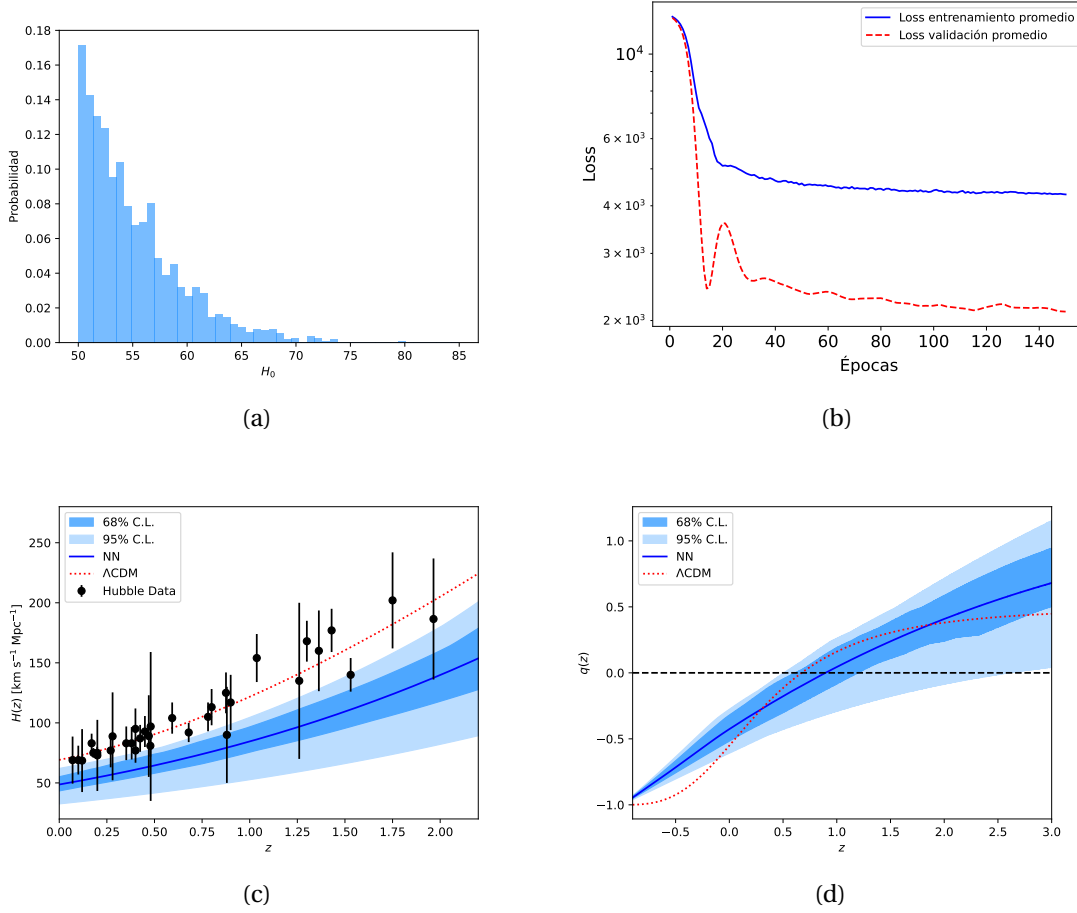


Figura 4.8: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.9. Modelo 1-150-250-1 $lr = 3 \times 10^{-3}$

En el modelo M9, se varió la tasa de aprendizaje, en este caso 3×10^{-3} . El valor obtenido de $H_0 = 68.14^{+1.50}_{-1.84} \text{ km/s/Mpc}$ es consistente con

los resultados de Planck, y los valores de $\chi^2 = 16.17$ y $MSE = 146.06$, los cuales nos dicen que el modelo ajusta bien a los datos de entrenamiento. En la Figura 4.9 muestra la reconstrucción de los parámetros cosmológicos de este modelo.

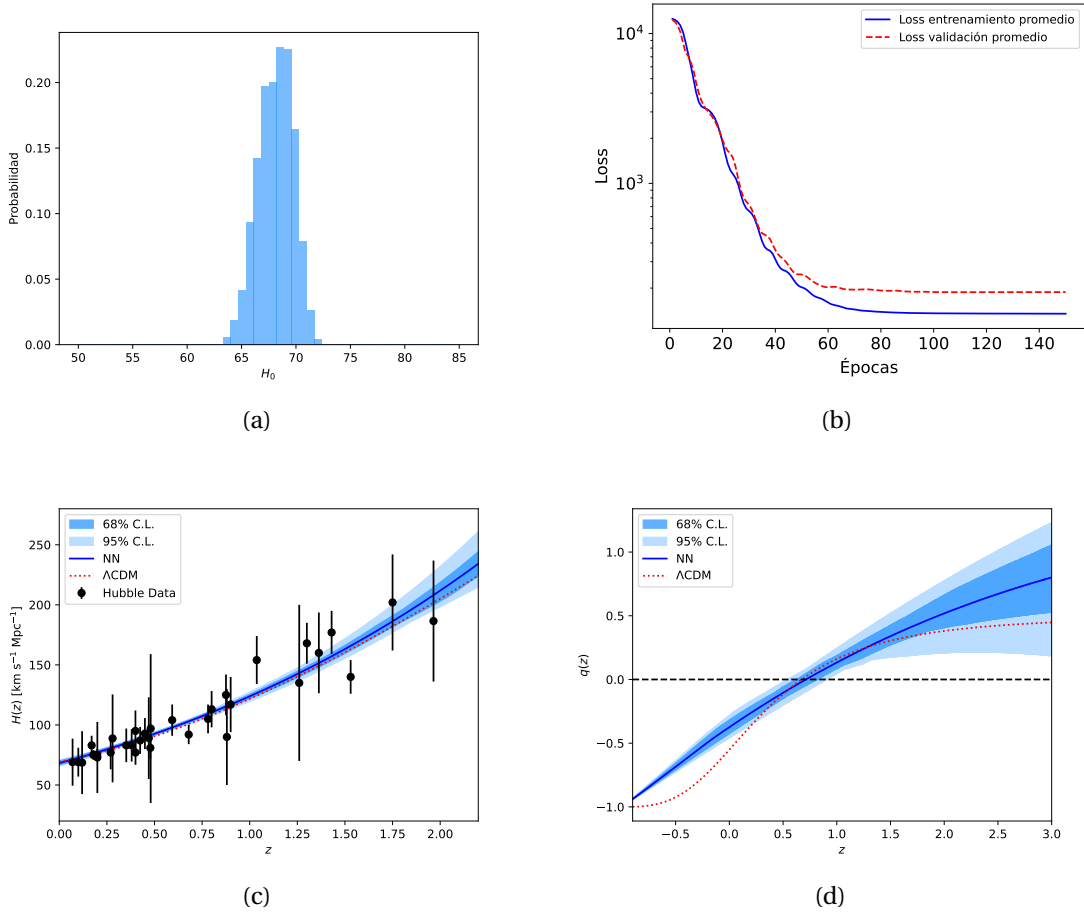


Figura 4.9: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

En la Figura 4.9(b), se observa que la función de pérdida disminuye durante todo el entrenamiento y sin presentar oscilaciones al final del entrenamiento. En el histograma de H_0 , Figura 4.9(a), podemos observar una distribución centrada en valores aceptados por la cosmología. En las Figuras 4.9(c) y 4.9(d), se pueden apreciar bandas de error pequeñas, además de que el valor promedio de la reconstrucción de $H(z)$ es cercano al valor de Λ CDM. En la reconstrucción de $q(z)$ se observa que

la transición de desaceleración a aceleración se realiza en $z_T = 0.71$, el cual es congruente con las observaciones actuales.

4.10. Modelo 1-150-250-1 $lr = 5 \times 10^{-3}$

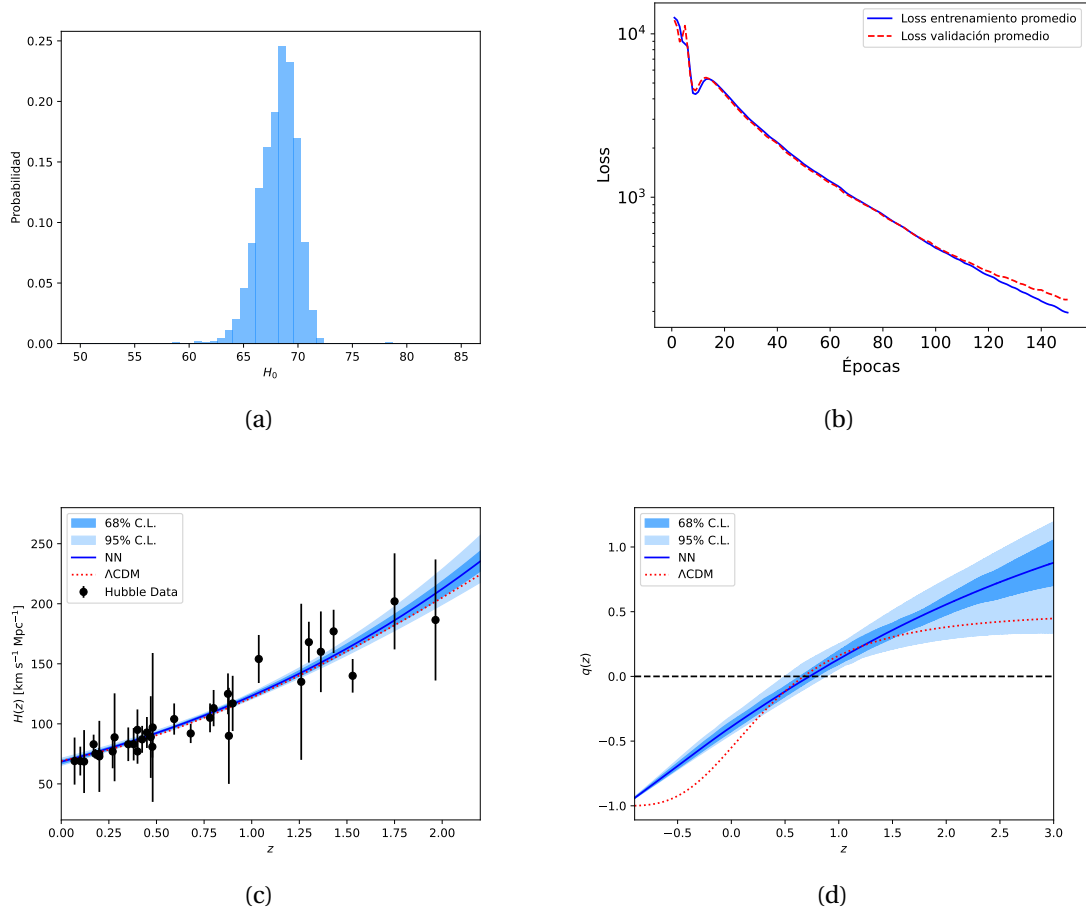


Figura 4.10: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

El modelo M10, entrenado con una tasa de aprendizaje de 5×10^{-3} , tiene la misma arquitectura que el modelo anterior pero con una tasa de aprendizaje de 5×10^{-3} . En comparación con el modelo M9, este muestra un peor desempeño, pues las métricas obtenidas por el modelo son $\chi^2 = 24.64$ y $MSE = 202.06$, las cuales son altas en comparación

con otros modelos con mejores resultados. Se obtuvo que el valor de $H_0 = 67.86^{+1.36}_{-2.18}$ km/s/Mpc es cercano al valor de Planck.

En la Figura 4.10(b), la función de pérdida decae durante todo el entrenamiento, aunque entre las épocas 10 y 20 se tiene un pico para después volver a decaer. Aunque en el histograma, Figura 4.10(a), se observa una distribución centrada pero más ancha. Las reconstrucciones de $H(z)$ y $q(z)$, en las Figuras 4.10(c) y 4.10(d), tienen un valor medio consistente con Λ CDM. Se obtuvo que la transición de $q(z)$ con un valor de $z_T = 0.72$, dentro del rango esperado, aunque con una ligera desviación respecto a los modelos con mejor ajuste.

4.11. Modelo 1-150-250-1 $lr = 8 \times 10^{-3}$

Para el modelo M11 se usó una tasa de aprendizaje de 8×10^{-3} . Para este caso se tuvo que el modelo no llegó a valores mínimos de las métricas, obteniendo valores de $\chi^2 = 547,64$ y $MSE = 3596$. Para el parámetro de Hubble se tiene el valor de $H_0 = 50.06^{+2.14}_{-66.57}$ km/s/Mpc el cual es muy bajo comparado con los valores de Planck y SH0ES. La Figura 4.11 muestra los resultados de esta arquitectura. En la Figura 4.11(b), se observa un comportamiento inestable de la función de pérdida, con picos al inicio del entrenamiento y una disminución lenta que termina el entrenamiento finalizando con valores altos. Las reconstrucciones de $H(z)$ y $q(z)$, en las Figuras 4.11(c) y 4.11(d), muestran barras de error muy grandes, en donde ninguna de las dos reconstrucciones muestra el comportamiento esperado, para el parámetro de desaceleración se tiene que la transición de desaceleración a aceleración se hace en $z_T = 0.36$, el cual no es consistente con ningún modelo cosmológico.

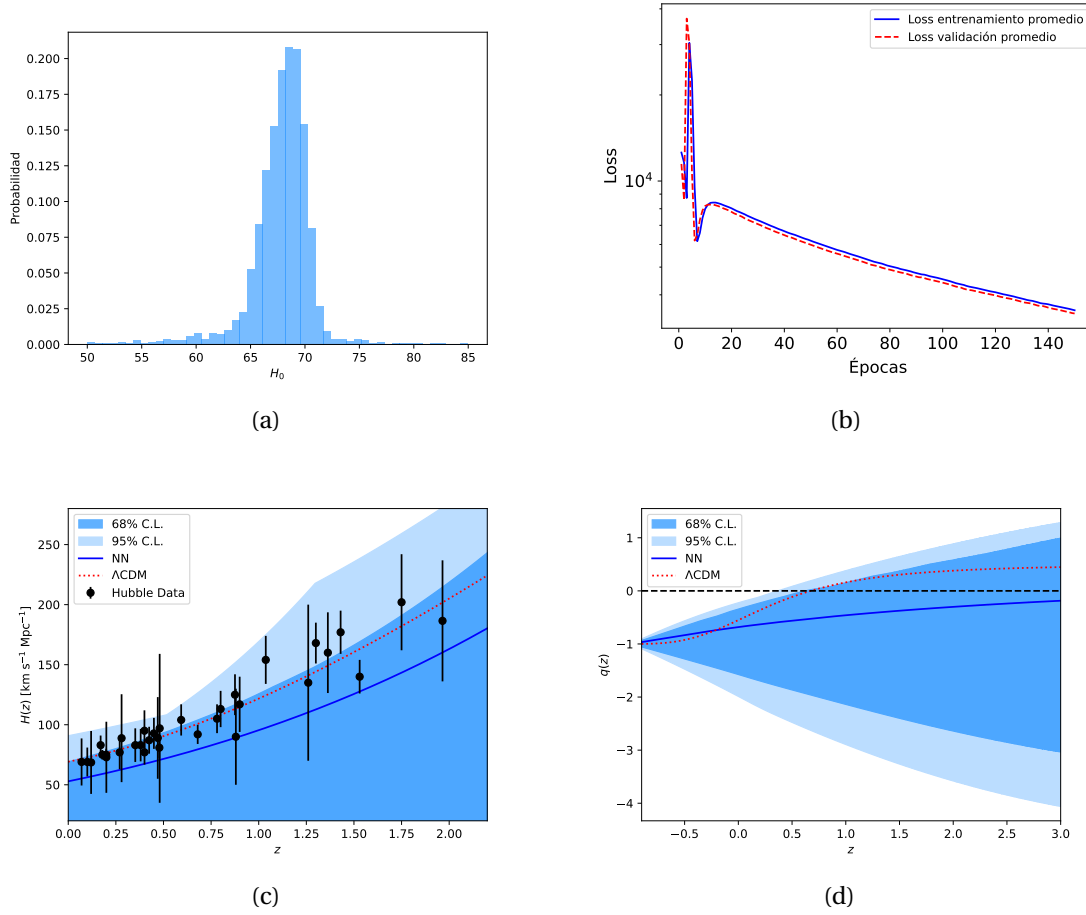


Figura 4.11: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.12. Modelo 1-10-20-1 150 épocas

Para el modelo M12, se usó una arquitectura más sencilla, usando 10 y 20 neuronas en las capas ocultas y se entrenó por 150 épocas. Se obtuvo un valor de $H_0 = 43.58^{+15.35}_{-20.00} \text{ km/s/Mpc}$ que se aleja de los valores de SH0ES y Planck. Para las métricas tenemos los valores $\chi^2 = 307.38$ y $MSE = 2076.72$, lo cual nos dice que el modelo no se entrena de forma correcta.

La función de pérdida, Figura 4.12(b), decrece sin oscilaciones, alcanzando valores bajos hacia el final del entrenamiento. El histograma de H_0 , Figura 4.12(a), muestra una distribución corrida hacia valores pe-

queños. La reconstrucción de $H(z)$, Figura 4.12(c), muestra una concavidad de la curva distinta al modelo Λ CDM, lo cual tiene implicaciones en la dinámica del universo. El parámetro $H(z)$ está relacionado con la tasa de expansión del universo; esto quiere decir que para la predicción de la red neuronal en el pasado, el universo tiene una expansión menos rápida a lo mostrado por el modelo estándar.

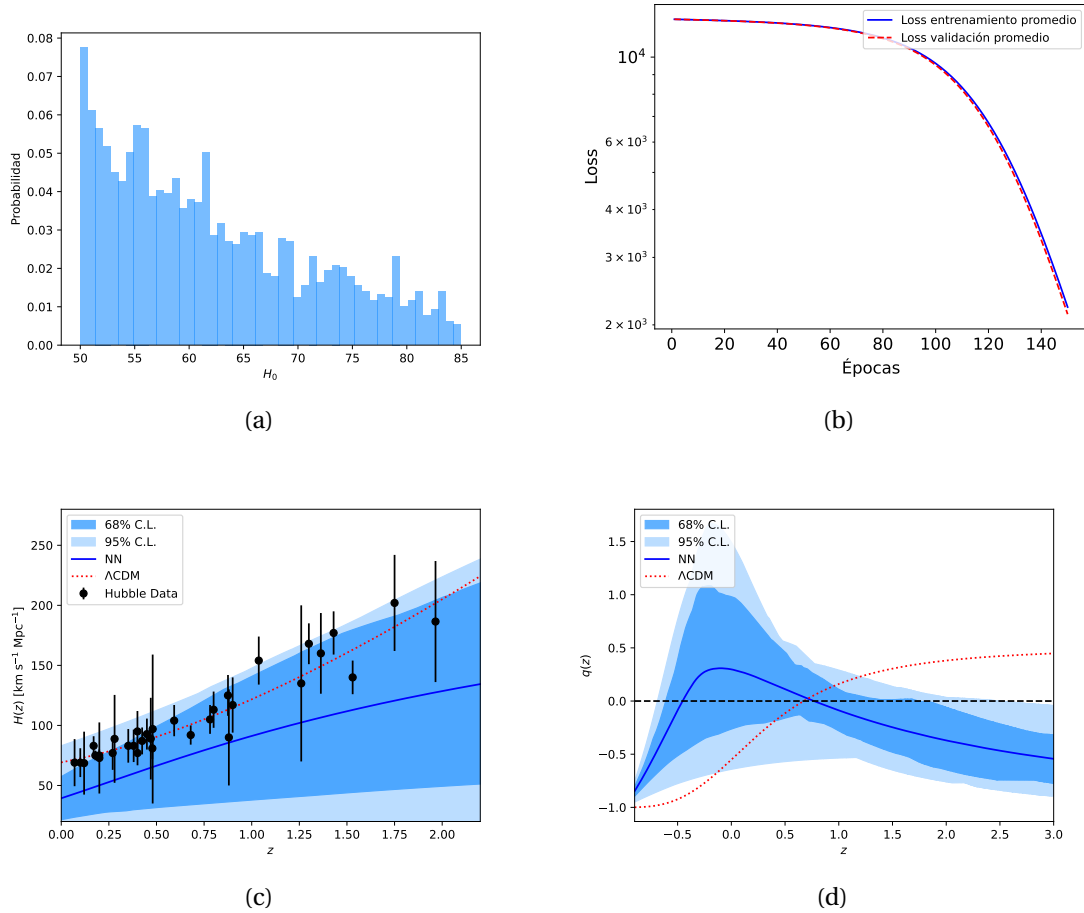


Figura 4.12: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

Para la reconstrucción de $q(z)$, Figura 4.12(d), muestra un comportamiento atípico, con un doble cruce en $q = 0$. Este fenómeno no es consistente con el modelo estándar de cosmología, que predice una única transición de $q(z)$ desde valores positivos (desaceleración) hacia negativos (aceleración). Además, nos dice que en el pasado el universo se en-

contraba en una fase de expansión acelerada, luego pasó a una fase de no expansión y finalmente, en el futuro, pasará nuevamente a una fase de aceleración.

4.13. Modelo 1-10-20-1 300 épocas

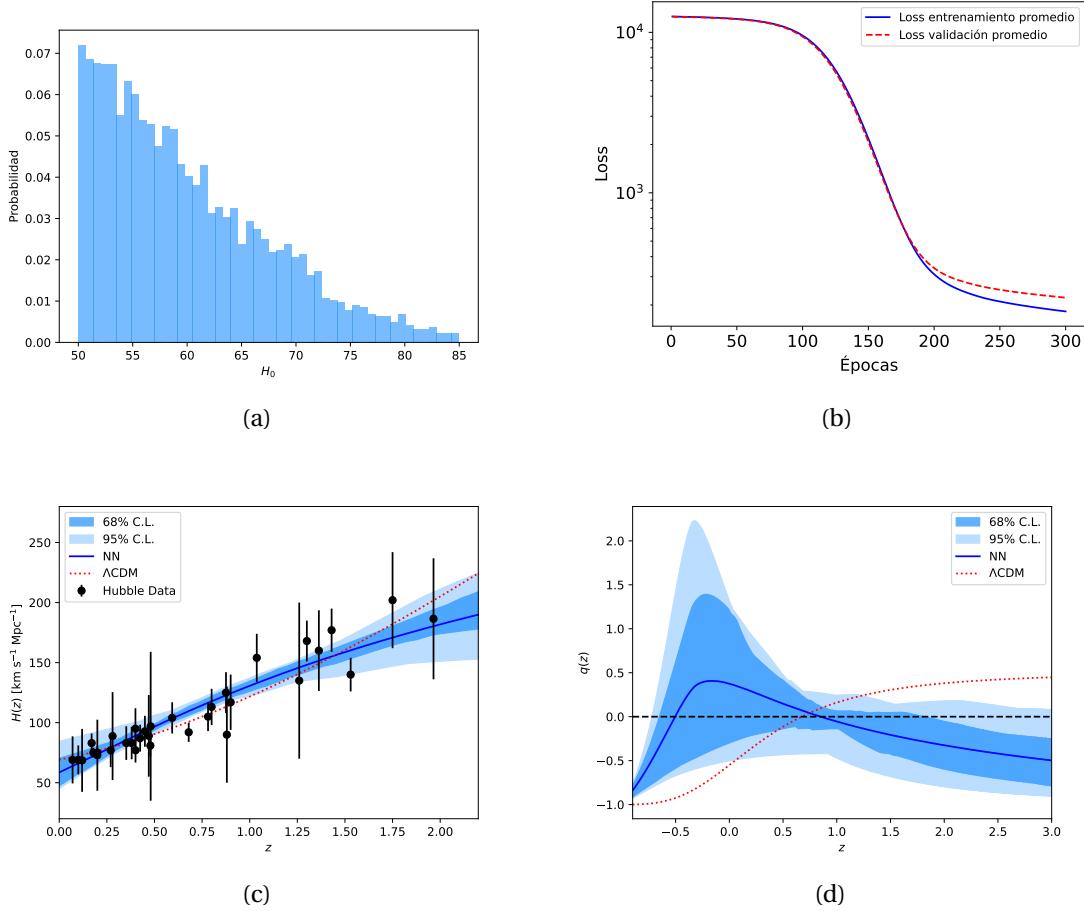


Figura 4.13: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

El modelo M13, con arquitectura 1-10-20-1 y 300 épocas para el entrenamiento, mostró una mejora respecto al modelo M12 en cuanto al descenso de la función de pérdida, aunque los resultados cosmológicos continúan mostrando desviaciones con el modelo estándar. El valor

obtenido de H_0 , y el valor de χ^2 indica que el modelo no se ajusta bien a los datos. La Figura 4.13 muestra los resultados de este modelo.

La función de pérdida, Figura 4.13(b), disminuye y alcanza un mínimo más bajo que en el modelo M12, lo cual nos dice que el modelo tiene una mejor convergencia del entrenamiento debido al aumento en el número de épocas. Sin embargo, el histograma de H_0 , Figura 4.13(a), presenta una distribución ancha y con tendencia a valores bajos, con muchos valores bajos. La reconstrucción de $H(z)$, Figura 4.13(c), mantiene una concavidad que no se presenta en el modelo Λ CDM, lo cual implica que la red neuronal sigue interpretando una expansión menos acelerada en el pasado.

En cuanto a la reconstrucción de $q(z)$, Figura 4.13(d), continúa un comportamiento atípico con un doble cruce en $q = 0$, similar al observado en M12. Esto sugiere nuevamente que el universo pasaría de una fase de aceleración a desaceleración y luego de vuelta a aceleración.

4.14. Modelo 1-10-20-1 500 épocas

El modelo M14, con arquitectura 1-10-20-1 y entrenado con 500 épocas. El valor obtenido de $H_0 = 56.07^{+5.76}_{-5.17}$ km/s/Mpc y el valor de $\chi^2 = 23.16$ indica que, a pesar del entrenamiento con más épocas, el ajuste sigue sin llegar a un valor óptimo. La Figura 4.14 muestra los resultados obtenidos para este modelo.

La función de pérdida, Figura 4.14(b), alcanza un mínimo ligeramente inferior al del modelo M13. El histograma de H_0 , Figura 4.14(a), conserva una distribución con tendencia a valores bajos. La reconstrucción de $H(z)$, Figura 4.14(c), mantiene una concavidad diferente del modelo Λ CDM, y con una expansión menos acelerada en épocas pasadas, a diferencia del comportamiento visto en el modelo estándar.

La reconstrucción del parámetro de desaceleración $q(z)$, Figura 4.14(d), vuelve a mostrar el patrón de doble cruce de la línea $q = 0$, como en los modelos M12 y M13. Este resultado significa que el modelo predice fases alternadas de expansión acelerada y desacelerada. Estos resultados nos dicen que el aumento de épocas de entrenamiento en una arquitectura más simple no corrige las deficiencias del modelo, y que la capacidad de predicción sigue estando restringida por el tamaño de la red.

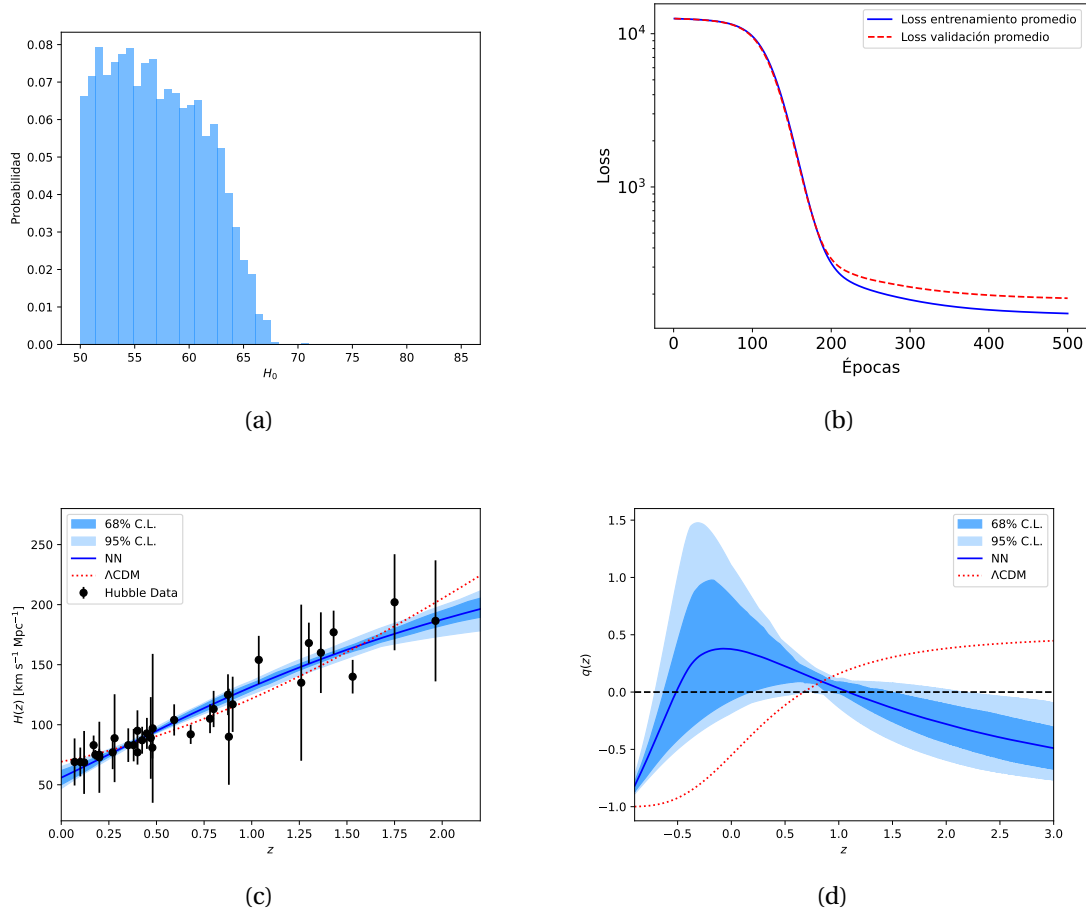


Figura 4.14: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.15. Modelo 1-150-50 150 épocas

El modelo M15, con arquitectura 1-150-50-1 y validación del 20%, muestra que el modelo no llega a un valor mínimo en las métricas de evaluación. El valor de H_0 se encuentra por debajo del valor de Λ CDM, con un valor de $H_0 = 52.68^{+11.11}_{-9.58}$, en donde podemos observar que los errores de la predicción aumentaron como se observa en la Figura 4.15(a).

En la Figura 4.15(b), se observa que la función de pérdida no llega a un valor mínimo, con un $MSE = 199.36$. Para la reconstrucción del parámetro de Hubble $H(z)$, Figura 4.15(c) se observa que al igual que los modelos anteriores se tiene una concavidad diferente al modelo estándar,

lo cual podría indicar que el universo presenta una tasa de expansión distinta.

La reconstrucción del parámetro de desaceleración $q(z)$, Figura 4.15(d), se muestra de nuevo un doble cruce en el eje z . Esto nos dice que el modelo predice que el universo ha estado alternando entre fases de aceleración y desaceleración.

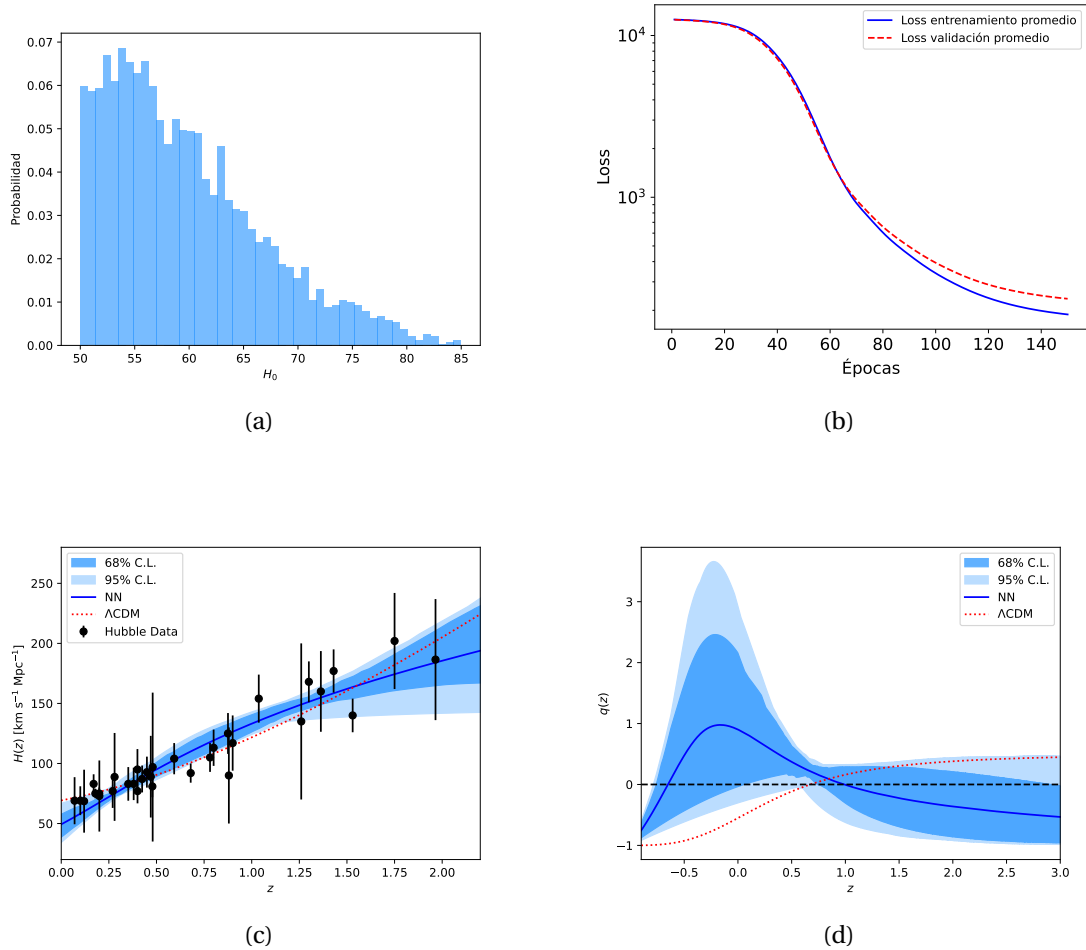


Figura 4.15: Resultado del entrenamiento de la red neuronal, superior derecha, imagen (a), histograma del valor de H_0 , superior izquierda, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior derecha, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior izquierda imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.16. Modelo 1-250-150-1 150 épocas

En el modelo M16, 1-250-150-1 se entrenó con 150 épocas, mejora en comparación con el modelo anterior que tiene menos neuronas

en sus capas ocultas. En la Figura 4.16(a), podemos observar que el histograma es más estrecho, pero preseta más predicciones hacia valores bajos en comparación con otros modelos. Para la función de pérdida, Figura 4.16(b), se percibe que la función de pérdida decae de forma suave y se vuelve estable en un mínimo.

En la reconstrucción del parámetro de Hubble, Figura 4.16(c), se observa que la reconstrucción de la predicción de la red neuronal es consistente con el modelo Λ CDM, teniendo la misma concavidad y estando la reconstrucción promedio cerca. Por último, para el parámetro de desaceleración, Figura 4.16(d), las predicciones se ajustan a la cosmología actual, mostrando una mayor banda de error para los valores altos de z .

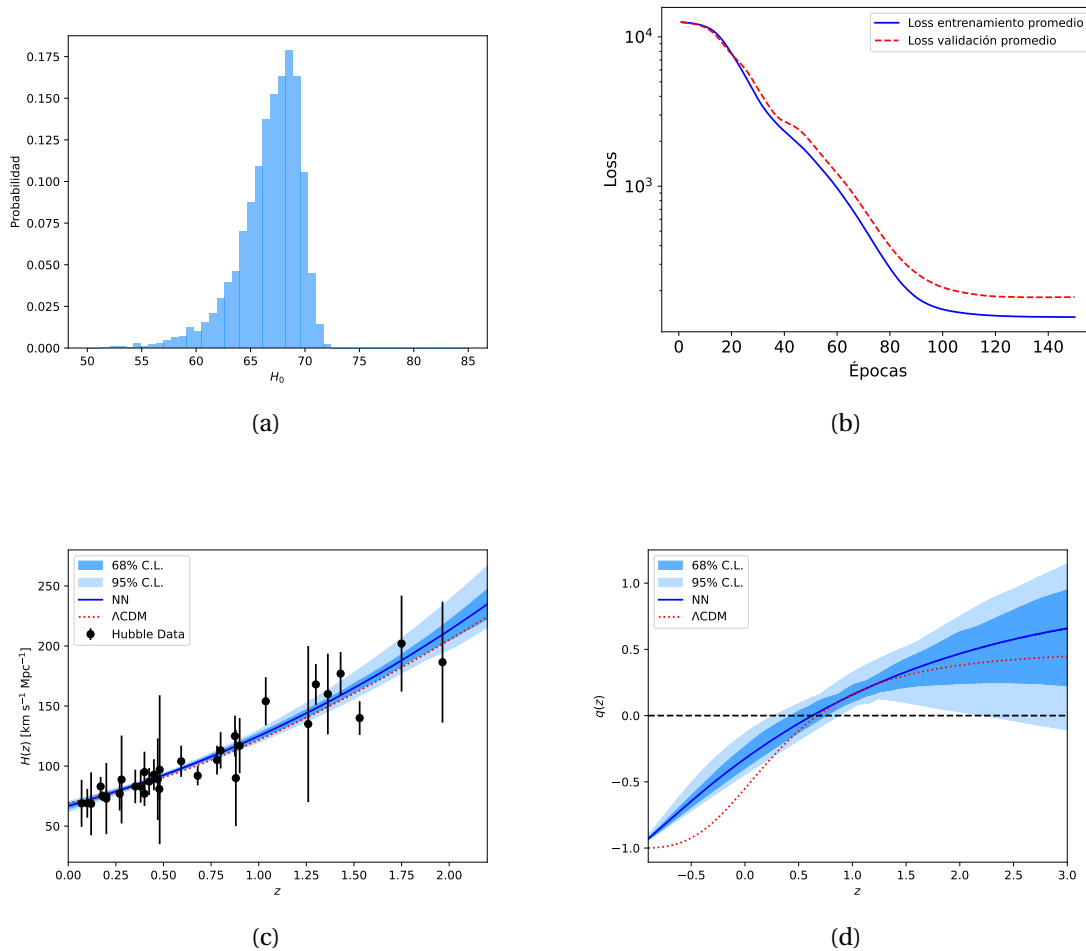


Figura 4.16: Resultado del entrenamiento de la red neuronal, superior derecha, imagen (a), histograma del valor de H_0 , superior izquierda, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior derecha, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior izquierda imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.17. Modelo 1-250-150-1 300 épocas

Para el modelo M17, se uso una red neuronal con 250 neuronas en la primera capa oculta y 150 neuronas en la segunda capa oculta, como el modeo anterior, solo que se entrenó por 300 épocas. En la Figura 4.17(a), podemos observar que el histograma es más ancho que el modelo anterior. En la Figura 4.17(b), se muestra que, el modelo ya no presenta mejora desde 100 épocas, lo cual nos dice que no diferencia en mejora con respecto al modelo M16. En la Figura 4.17(c), se aprecia que la reconstrucción del parámetro de desaceleración se acerca más el modelo Λ CDM, pero presenta una banda de error más grande al final de la curva.

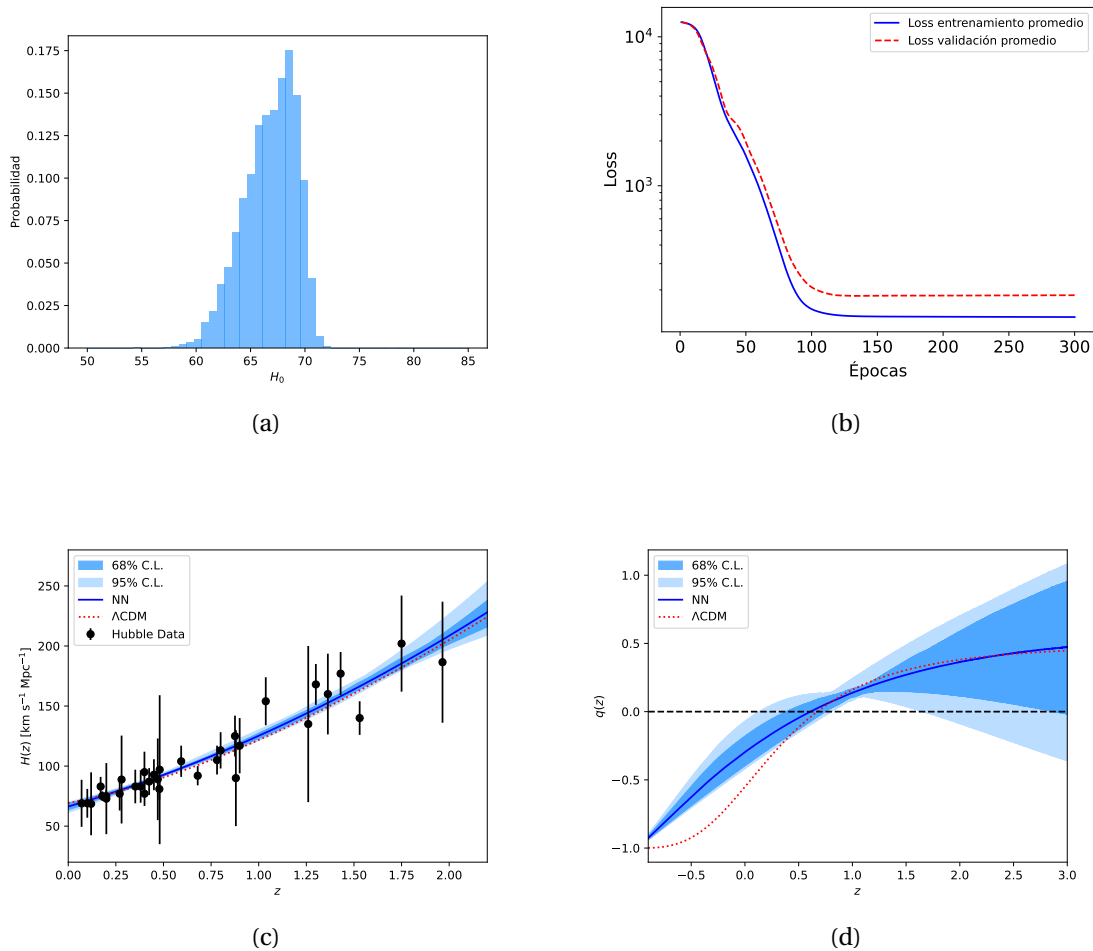


Figura 4.17: Resultado del entrenamiento de la red neuronal, superior derecha, imagen (a), histograma del valor de H_0 , superior izquierda, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior derecha, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior izquierda imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

4.18. Modelo 1-250-150-1 *dropout* 20% 300 épocas

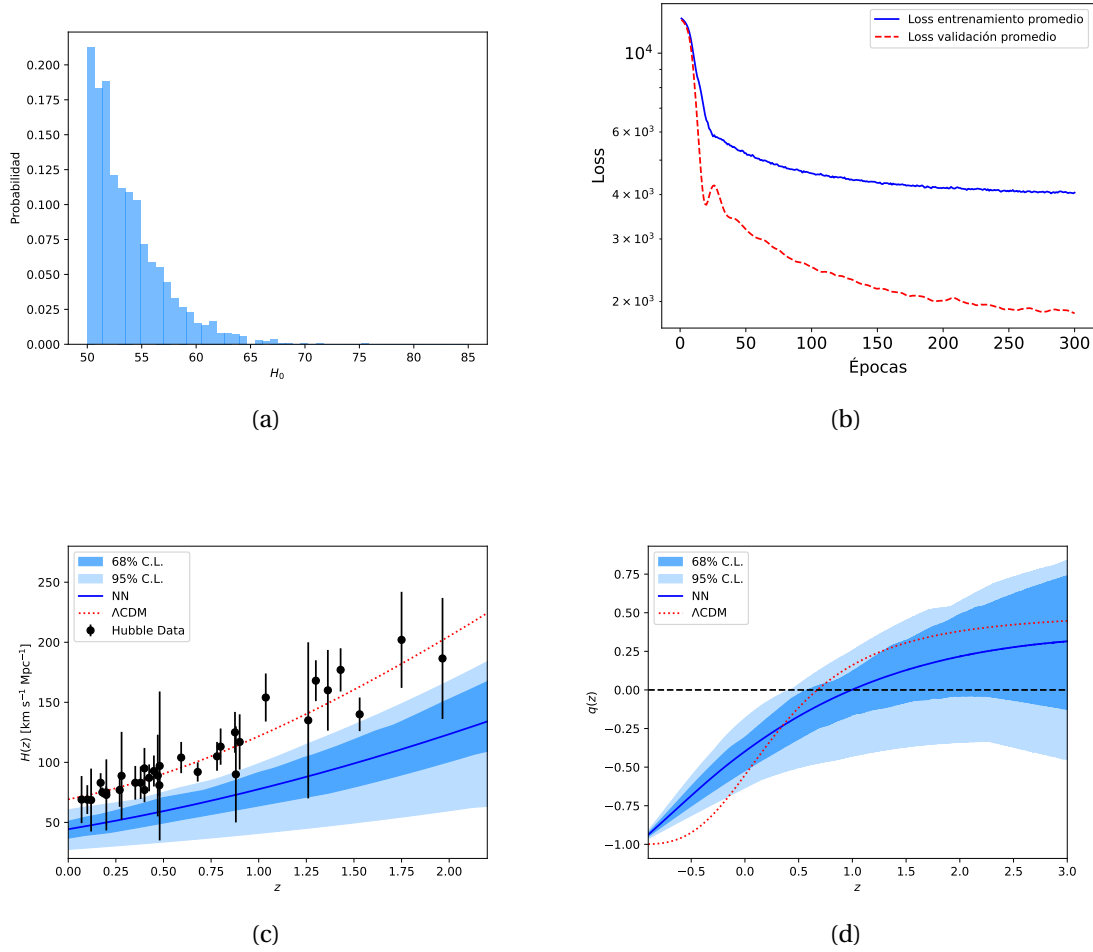


Figura 4.18: Resultado del entrenamiento de la red neuronal, superior izquierda, imagen (a), histograma del valor de H_0 , superior derecha, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior izquierda, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior derecha imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

En el modelo M18, se mantiene la misma arquitectura que en el modelo M17, pero se agrega el uso de *dropout* del 20%. En la Figura 4.18(a), se observa que el histograma se encuentra muy desplazado hacia valores bajos de H_0 . En la Figura 4.18(b), podemos observar la función de pérdida, en donde vemos una diferencia entre las curvas del entrenamiento y la validación, esta diferencia puede ser causada por el uso de *dropout*, ya que dificulta que el modelo aprenda.

Para la reconstrucción del $H(z)$, Figura 4.18(c), las predicciones se encuentran por debajo del valor de ΛCDM , y los intervalos de con-

finza se ven más anchos. Por último, en la Figura 4.18(d), se muestra el parámetro de desaceleración, en donde se observa que la transición de desaceleración a aceleración se hace antes que en el modelo Λ CDM.

4.19. Modelo 1-1024-512-1 150 épocas

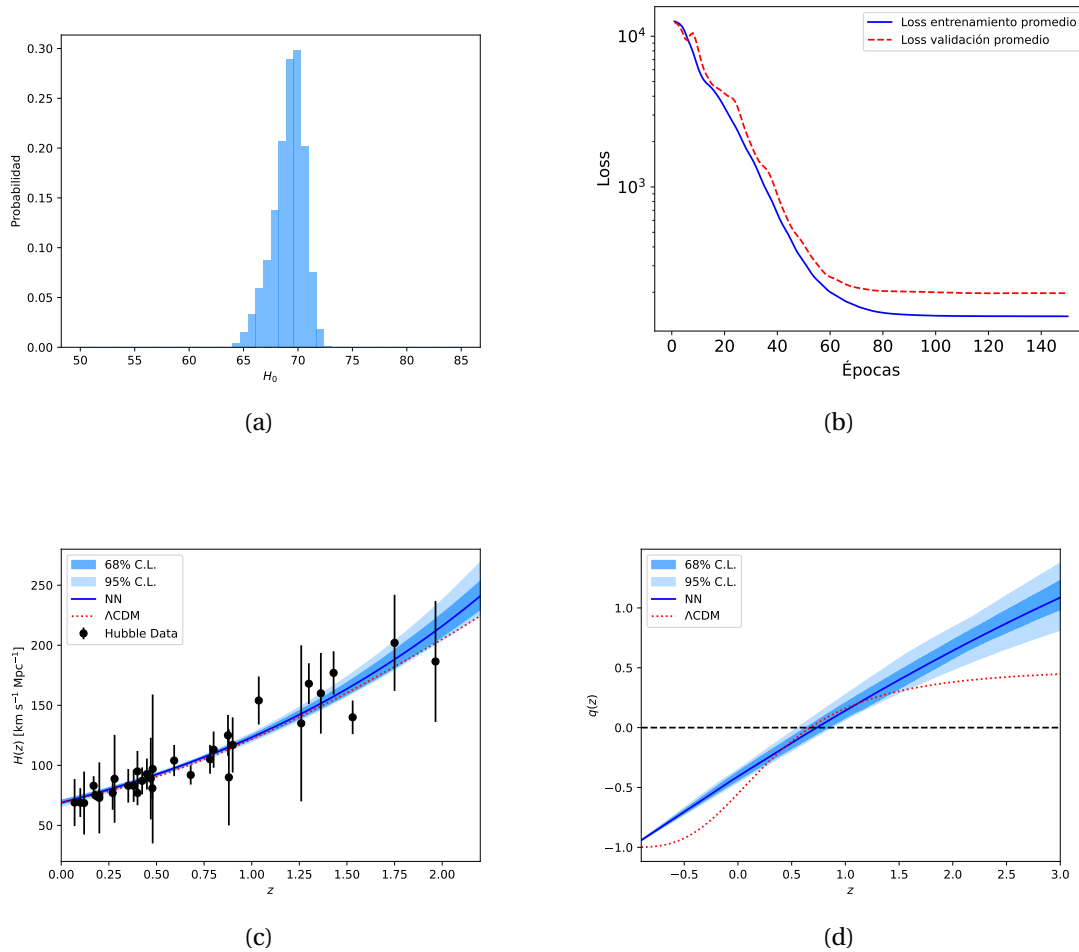


Figura 4.19: Resultado del entrenamiento de la red neuronal, superior derecha, imagen (a), histograma del valor de H_0 , superior izquierda, imagen (b) gráfica de función de pérdida durante el entrenamiento de la red neuronal, inferior derecha, imagen (c) reconstrucción del parámetro de Hubble mediante la red neuronal, inferior izquierda imagen (d) reconstrucción de parámetro de desaceleración con red neuronal.

Para el modelo M19, se tiene una red con más neuronas en sus capas ocultas, con 1024 y 512 neuronas respectivamente. En la Figura 4.19(a), se muestra el histograma de los valores predichos para H_0 , en donde se observa que no hay una dispersión grande, lo que se puede observar también en la función de pérdida, 4.19(b), se observa que ambas

curvas se encuentran cercanas, lo que indica que no hay sobreajuste. En la reconstrucción del parámetro de Hubble, 4.19(c), se observa que las bandas de error son pequeñas en todo el rango de z . Y finalmente, para $q(z)$, Figura 4.19(d), el cruce que indica la aceleración en la expansión del universo se hace muy cerca de Λ CDM. En conjunto, estos resultados indican que este modelo muestra buenos resultados comparables con los mejores modelos entrenados, pero presenta un tiempo de entrenamiento 6 veces mayor que modelos con los que son comparables estos resultados.

4.20. Tensión de Hubble

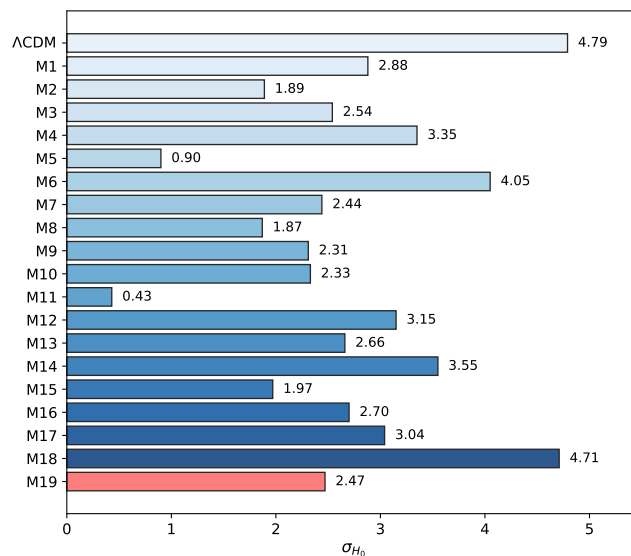


Figura 4.20: Valores obtenidos para σ con respecto a los valores de SH0ES y Planck

En la Tabla 4.2 y la Figura 4.20 se muestran los valores de σ calculados para los modelos de redes neuronales en comparación con los valores reportados por SH0ES y Planck. El valor de σ indica qué tan probable es que los valores ambas pertenezcan a la misma distribución, es decir, mientras más elevado sea el valor de σ menor es la probabilidad. En la tabla, podemos observar que todos los modelos de redes neuronales disminuyen este valor, lo que nos dice que todos estos modelos disminuyen la tensión de Hubble.

Los valores de σ en relación con SH0ES muestran que los modelos presentan una gran dispersión. El modelo M5, que utiliza un conjunto

Modelo	Nombre	σ_{SH0ES}	σ_{Planck}
ΛCDM	-	4.79	0.39
1-50-150-1	M1	2.88	0.08
1-100-300-1	M2	1.89	0.39
1-150-250-1	M3	2.54	0.57
1-50-150-1, val 20 %	M4	3.35	1.13
Entrenamiento 100 %, validación 100 %	M5	0.90	0.15
Dropout 20 %	M6	4.05	3.28
Dropout 10 %, val 20 %	M7	2.44	1.64
Dropout 20 %, 1-100-300-1	M8	1.87	1.25
$lr = 3 \times 10^{-3}$	M9	2.31	0.38
$lr = 5 \times 10^{-3}$	M10	2.33	0.07
$lr = 8 \times 10^{-3}$	M11	0.43	0.34
1-10-20-1 150 épocas	M12	3.15	2.05
1-10-20-1 300 épocas	M13	2.66	1.71
1-10-20-1 500 épocas	M14	3.55	2.40
1-150-50-1	M15	1.97	1.43
1-250-150-1, 150 épocas	M16	2.70	0.25
1-250-150-1, 300 épocas	M17	3.04	0.28
1-250-150-1, 300 épocas, dropout 20 %	M18	4.71	3.70
1-1024-512-1, 150 épocas	M19	2.47	1.62
Promedio ponderado	-	4.38	1.48

Tabla 4.2: Valores de σ obtenidos para cada modelo de red neuronal. La columna σ_{SH0ES} representa la discrepancia con los datos de SH0ES, y σ_{Planck} con los datos de Planck.

de entrenamiento y validación del 100 %, tiene una mayor compatibilidad con los datos de SH0ES. Por el contrario, el modelo M6, que emplea *dropout* del 20 %, tiene una mayor discrepancia.

En los valores obtenidos comparados con Planck, se observa una menor dispersión en los valores de σ . El modelo M2 presenta el menor valor de σ de los modelos que presentan el mejor ajuste.

4.21. Discusión de resultados

El objetivo de este trabajo es estimar los parámetros cosmográficos mediante redes neuronales y analizar la sensibilidad de los hiperparámetros del modelo. La Tabla 4.3 muestra un resumen de los resultados obtenidos de las distintas configuraciones, las cuales varían en el número de neuronas por capa, la inclusión de técnicas de regularización, el uso de conjuntos de validación y la tasa de aprendizaje utilizada durante el entrenamiento. Para la comparación de los valores reconstruidos de H_0 , q_0 y z_T , la Figura 4.21 muestra los resultados de cada modelo junto con los valores reportados por Planck y SH0ES.

Nombre	χ^2	MSE	H_0	q_0	z_T
Λ CDM	15.24	140.93	$67.66^{+0.42}_{-0.42}$	$-0.53^{+0.03}_{-0.01}$	$0.63^{+0.11}_{-0.11}$
M1	16.35	142.30	$67.25^{+1.72}_{-0.57}$	$-0.33^{+0.08}_{-0.05}$	$0.67^{+0.05}_{-0.11}$
M2	17.80	155.03	$68.32^{+2.27}_{-1.83}$	$-0.37^{+0.05}_{-0.03}$	$0.72^{+0.04}_{-0.06}$
M3	17.47	152.95	$68.33^{+1.53}_{-1.45}$	$-0.38^{+0.05}_{-0.03}$	$0.72^{+0.06}_{-0.06}$
M4	16.19	146.24	$68.19^{+1.34}_{-1.76}$	$-0.37^{+0.05}_{-0.04}$	$0.71^{+0.07}_{-0.12}$
M5	16.16	145.82	$68.16^{+1.30}_{-1.97}$	$-0.37^{+0.05}_{-0.04}$	$0.70^{+0.07}_{-0.12}$
M6	288.14	2096.76	$44.25^{+7.02}_{-6.43}$	$-0.40^{+0.11}_{-0.10}$	$0.95^{+0.49}_{-0.30}$
M7	111.21	883.93	$55.80^{+6.66}_{-10.43}$	$-0.40^{+0.09}_{-0.08}$	$0.89^{+0.34}_{-0.22}$
M8	294.93	2147.24	$44.84^{+7.76}_{-7.99}$	$-0.42^{+0.10}_{-0.10}$	$0.99^{+0.48}_{-0.29}$
M9	16.17	146.06	$68.14^{+1.50}_{-1.84}$	$-0.37^{+0.05}_{-0.04}$	$0.71^{+0.07}_{-0.12}$
M10	33.24	258.94	$67.57^{+1.62}_{-2.10}$	$-0.37^{+0.05}_{-0.04}$	$0.71^{+0.07}_{-0.12}$
M11	547.64	3596.10	$50.06^{+2.14}_{-66.57}$	$-0.70^{+0.07}_{-1.08}$	$0.36^{+0.12}_{-1.58}$
M12	307.38	2076.72	$43.58^{+15.35}_{-20.00}$	$0.35^{+0.69}_{-0.68}$	$1.01^{+0.73}_{-0.89}$
M13	28.69	190.40	$57.06^{+9.42}_{-6.63}$	$0.41^{+0.57}_{-0.52}$	$0.77^{+0.39}_{-0.94}$
M14	23.16	157.17	$56.07^{+5.76}_{-5.17}$	$0.34^{+0.42}_{-0.35}$	$0.86^{+0.36}_{-1.11}$
M15	34.57	199.36	$52.68^{+11.11}_{-9.58}$	$0.58^{+1.01}_{-0.59}$	$0.80^{+0.57}_{-1.05}$
M16	16.29	143.36	$66.77^{+1.91}_{-3.18}$	$-0.31^{+0.12}_{-0.05}$	$0.74^{+0.11}_{-0.17}$
M17	16.16	142.71	$66.76^{+2.16}_{-2.46}$	$-0.31^{+0.12}_{-0.06}$	$0.77^{+0.11}_{-0.18}$
M18	247.05	1843.91	$45.54^{+4.40}_{-7.44}$	$-0.37^{+0.13}_{-0.10}$	$1.02^{+0.57}_{-0.34}$
M19	16.51	150.89	$69.11^{+1.07}_{-1.33}$	$-0.41^{+0.04}_{-0.02}$	$0.75^{+0.05}_{-0.12}$
Promedio ponderado	16.52	147.01	$68.00^{+0.49}_{-0.49}$	$-0.36^{+0.01}_{-0.01}$	$0.71^{+0.02}_{-0.02}$

Tabla 4.3: Reconstrucción de parámetros cosmográficos mediante distintas configuraciones de red neuronal. Se reportan los valores de χ^2 , error cuadrático medio (MSE), parámetro de Hubble H_0 , parámetro de desaceleración q_0 y corrimiento al rojo de transición z_T .

Además de los resultados por modelo, se incluyó en la Tabla 4.3 el promedio ponderado de los parámetros cosmográficos estimados. Este promedio se calculó considerando únicamente los modelos que presen-

taron un buen ajuste estadístico, en términos de χ^2 y de MSE. Se utilizaron los modelos M1, M2, M3, M4, M5, M9, M16, M17 y M19. El promedio ponderado fue obtenido asignando a cada modelo un peso inversamente proporcional a su error en cada uno de los parámetros, de forma que los modelos con mejor ajuste contribuyeran más a la media final.

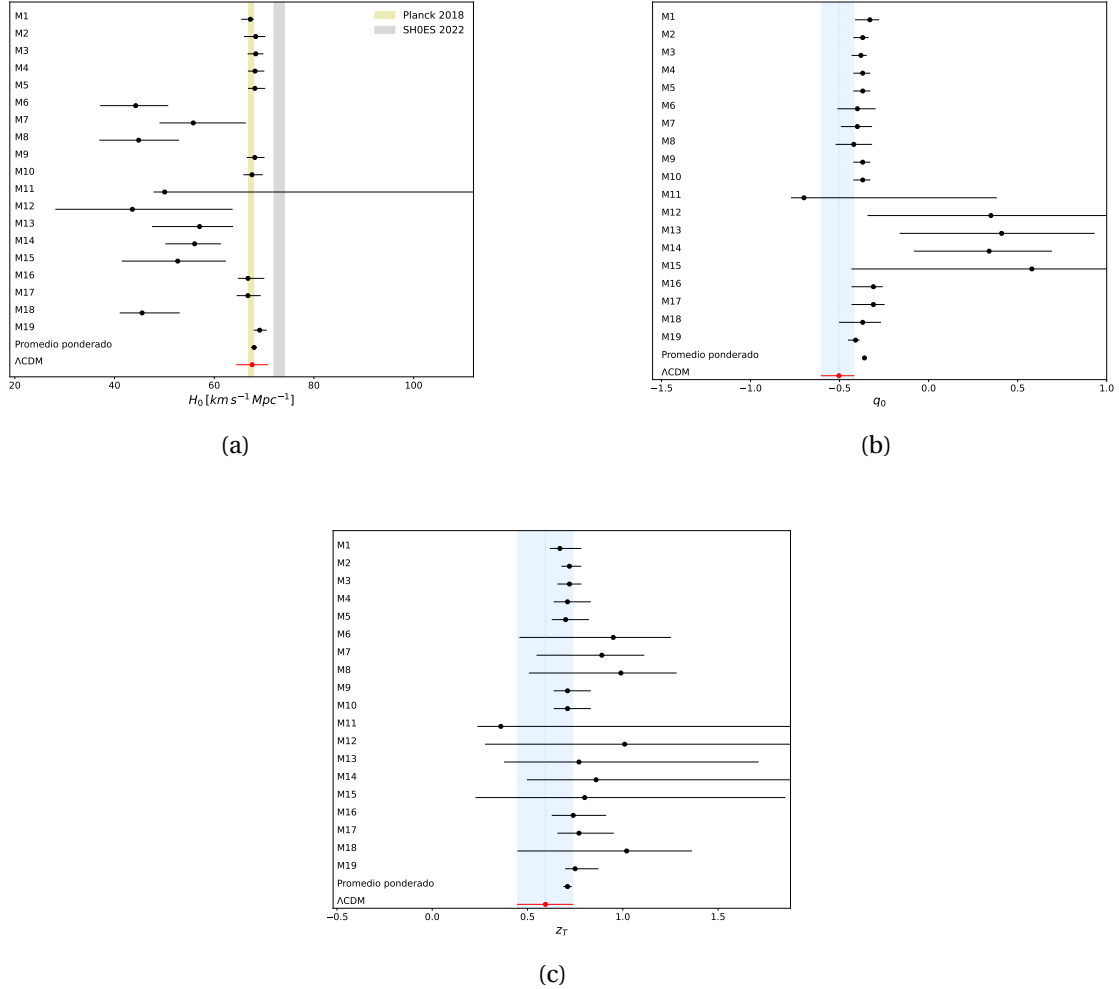


Figura 4.21: Comparación de los valores obtenidos por cada uno de los modelos de redes neuronales con modelos cosmológicos. Figura (a), superior izquierda comparación de valores de H_0 , figura (b) superior derecha valores de q_0 , figura (c) inferior comparación de valores de z_T .

Para evaluar la precisión de las estimaciones, se compararon los valores obtenidos con el modelo cosmológico estándar ΛCDM . Se observa que las configuraciones M1 y M4 presentan los valores más bajos de χ^2 y MSE, lo que nos dice que es un mejor ajuste de estas arquitecturas en comparación con las demás configuraciones. Estos modelos, que no incluyen técnicas de regularización como *dropout*, dan valores de H_0 , q_0

y z_T cercanos a los valores esperados, con lo que concluimos que la red neuronal es capaz de reconstruir los parámetros cosmográficos de manera satisfactoria.

Una de las características importantes en la reconstrucción de parámetros cosmográficos mediante redes neuronales es la sensibilidad de las estimaciones a las características estructurales de la red, incluyendo la cantidad de capas ocultas, el número de neuronas en cada capa y el uso de técnicas de regularización. En este estudio, se encontró que la inclusión de *dropout* afecta la precisión de las predicciones. Se observa que los modelos M6, M7 y M8, que usan el *dropout* tienen un mal ajuste de los parámetros, que podemos observar por el aumento en los valores de χ^2 y MSE, así como por la reducción del valor de H_0 y el incremento en sus bandas de error. Estos resultados indican que el dropout en lugar de mejorar el entrenamiento del modelo, limita la capacidad de la red neuronal para ajustar correctamente los datos observacionales y los parámetros cosmográficos. Aunque el *dropout* es útil para evitar el sobreajuste en problemas con grandes cantidades de datos, en este caso no ha dado buenos resultados debido a la cantidad limitada de datos.

El modelo M6, que tiene mayor *dropout*, muestra el mayor error cuadrático medio y las desviaciones más grandes en q_0 y z_T , lo que indica que una regularización alta puede sesgar las estimaciones y reducir la precisión de la reconstrucción de parámetros cosmográficos. Por otro lado, el modelo M4, que incluye un conjunto de validación del 20 %, tiene un buen ajuste. Las predicciones de H_0 , q_0 y z_T son cercanas a los valores esperados.

Además, se realizaron pruebas modificando la tasa de aprendizaje. El modelo M9, con una tasa de aprendizaje $lr = 3 \times 10^{-3}$, tiene un comportamiento similar al de los mejores modelos, con un valor de χ^2 bajo y un valor estimado de H_0 consistente con el modelo Λ CDM. En cambio, los modelos M10 y M11, con tasas de aprendizaje más elevadas, no logran entrenarse de la forma esperada y las predicciones no son buenas. El modelo M11 con $lr = 8 \times 10^{-3}$, alcanzó valores muy altos de χ^2 y MSE, y una estimación muy mala de H_0 . Esto nos dice que una tasa de aprendizaje alta puede llevar a que el modelo no aprenda de forma adecuada.

Se realizaron pruebas con redes neuronales con arquitectura más simple como en los modelos M12, M13 y M14, que utilizan pocas neu-

ronas (1-10-20-1) y fueron entrenados con 150, 300 y 500 épocas, respectivamente. Se observó que, con pocas neuronas, la red tiene una capacidad limitada para representar con precisión la relación entre los datos observacionales y los parámetros cosmográficos. El modelo M12, con 150 épocas, presenta una alta incertidumbre en las estimaciones y un valor de q_0 positivo, lo que se contradice con el modelo estándar. Si se aumenta el número de épocas, como en M13 y M14, se observa una mejora en el ajuste de MSE y χ^2 . Además, estos modelos, en el parámetro de desaceleración, presentan un doble cruce en el eje x, lo que indica dos transiciones entre expansión acelerada y desacelerada del universo, lo cual no es consistente con el modelo Λ CDM, ni con ningún modelo cosmológico estudiado en el mismo rango de redshift.

Los modelos con arquitecturas más sencillas, sin validación y sin *dropout*, presentan mayor dispersión en las distribuciones de H_0 , pero se mantienen cercanos a los valores de los modelos Λ CDM. Esta dispersión en los histogramas indican que el modelo realiza predicciones en un rango más amplio, lo que reduce la precisión. Esto también tiene impacto en $H(z)$, en donde el intervalo de confianza se hace más amplio y la función de pérdida presenta oscilaciones, lo que indica que el modelo puede no estar capturando de manera correcta los patrones, probablemente por la falta de un conjunto de validación.

Las redes con mayor número de neuronas en las capas ocultas, entre 150-300 neuronas, realizan histogramas de H_0 más estrechas, lo que nos dice que el modelo realiza mejores predicciones. En estos casos es importante incluir un conjunto de validación para evitar que el modelo realice sobreajuste, sobre todo por la poca cantidad de datos con la que se cuenta.

Finalmente se realizaron pruebas invirtiendo el orden de el número de neuronas en las capas ocultas, como en los modelos M4 (250-150) y M16(150-250). Los resultados muestran que la diferencia entre estos modelos no es significativa como se muestra en la función de pérdida MSE y la reconstrucción de los parámetros cosmográficos para ambos modelos. La poca diferencia entre los modelos podría deberse a la cantidad limitada de datos para el entrenamiento. Para el modelo M19, el cual tiene 1024 y 512 neuronas en sus capas ocultas, tiene resultados comparables a los mejores modelos, como que se agregó un conjunto de validación, lo que permitió que no hiciera sobreajuste, aunque compu-

tacionalmente no es efectivo, pues el tiempo de computo es 3 veces mayor a modelos con arquitecturas más sencillas como M4.

5

Conclusiones

Con este trabajo se demuestra que las redes neuronales son útiles para encontrar patrones que a simple vista no se podrían ver aún con conjuntos de datos pequeños y el uso de estos modelos de inteligencia se pueden aplicar a la cosmología para realizar predicciones de los parámetros cosmográficos. Además de las predicciones realizadas por cada modelo, se incluyó el promedio ponderado de los parámetros cosmográficos estimados. Este promedio se calculó considerando únicamente los modelos que presentaron un buen ajuste estadístico, en términos de χ^2 y de MSE. Se utilizaron los modelos M1, M2, M3, M4 y M9. El promedio ponderado fue obtenido asignando a cada modelo un peso inversamente proporcional a su error en cada uno de los parámetros, de forma que los modelos con mejor ajuste contribuyeran más a la media final.

- Aunque el *dropout* es una técnica para evitar el sobreajuste, en este estudio no se obtuvieron buenos resultados ya que se contaba con una cantidad limitada de datos y la necesidad de mantener estructuras simples.
 - Dividir la muestra y usar un conjunto de validación mejoró la precisión de los modelos, ya que es conjunto independiente para evaluar el rendimiento del modelo y ajustar los pesos y a evitar que el modelo se adapte demasiado a los datos de entrenamiento.
 - Los valores de los parámetros cosmográficos de los modelos con mejores resultados, (M1, M2, M3, M4 y M9), así como los del promedio ponderado tienden a los valores obtenidos por el telescopio Planck.
-

-
- Con las predicciones realizadas por los modelos de redes neuronales, se obtuvo un valor de la constante de Hubble de $H_0 = 68.19^{+0.57}_{-0.57}$ km/s/Mpc, calculada mediante el promedio ponderado de los modelos con mejor desempeño. Este resultado muestra que el valor se encuentra a 1.3σ del valor de Planck y a 3.6σ de SH0ES, lo que podría indicar que la red neuronal podría estar captando patrones en los datos que métodos cosmológicos tradicionales no son capaces de identificar.
 - La cantidad limitada de datos disponibles fue un factor que afecta el rendimiento del modelo. Ya que las redes neuronales son sensibles a la cantidad de datos, se usaron modelos sencillos para evitar el sobreajuste.
 - Los valores MSE y χ^2 nos dicen qué tan bien el modelo se ajusta a los datos:
 - Modelos sin dropout, (M1, M2, M3, M4 y M9), tienen valores bajos MSE y χ^2 , es decir, menores o del orden de magnitud de los obtenidos con el modelo Λ CDM lo que indica un buen ajuste a los datos.
 - Modelos con dropout (M6, M7, M8), con tasas de aprendizaje superiores a 3×10^{-3} y redes con pocas neuronas en las capas ocultas (M12, M13, M14) presentan MSE y χ^2 altos en comparación con los obtenidos con Λ CDM, indicando que el modelo no ajusta bien.
 - Se probaron modelos con diferente número de épocas en la red neuronal con 10-20 neuronas en las capas ocultas, M12-M14. Aunque el aumento de épocas mejora el ajuste, también incrementa el error en los parámetros H_0 , q_0 y z_T . Además de que en el parámetro q_0 se tiene doble cruce en el eje del corrimiento al rojo z , lo que indica que el universo ha pasado de aceleración a desaceleración en más de una ocasión, siendo este comportamiento contradictorio con los datos observacionales y no presentes en los modelos cosmológicos tradicionales.
 - Las pruebas con redes neuronales de arquitectura reducida, modelos M12-M14, demostraron que son eficientes computacionalmen-

te, pero no tienen la capacidad suficiente para modelar relaciones cosmológicas, por lo que para este trabajo no fueron útiles.

- El orden del número de neuronas en las capas ocultas parecen no tener un impacto significativo en el entrenamiento de la red neuronal y basado en las gráficas de la función de pérdida se puede observar que no existe sobre ajuste cuando se tienen más neuronas en la segunda capa oculta.
- Una arquitectura con más neuronas, M19, logra resultados comparables con los modelos más simples, pero no es viable por su alto costo computacional.

En conclusión, este trabajo muestra que las redes neuronales son una herramienta eficaz para la reconstrucción de parámetros cosmológicos a partir de datos observacionales. A través de un análisis de diferentes arquitecturas, técnicas de regularización y variaciones en los hiperparámetros, se tienen configuraciones que permiten obtener buenas estimaciones de H_0 , q_0 y z_T sin hacer uso de modelos cosmológicos en los cuales se pueden incluir sesgos. Los resultados indican que, aún con conjuntos de datos pequeños, es posible obtener reconstrucciones precisas usando modelos de inteligencia artificial.

Bibliografía

- [1] Steven Weinberg. *Cosmology*. Oxford University Press, 2008. 14
 - [2] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. Academic Press, 2nd edition, 2020. 14
 - [3] Wendy L. Freedman and Barry F. Madore. The hubble constant. *Annual Review of Astronomy and Astrophysics*, 48:673–710, 2010. 14
 - [4] Sherry H. Suyu, Tommaso Treu, S. Hilbert, et al. Two accurate time-delay distances from strong lensing: Implications for cosmology. *The Astrophysical Journal*, 766(1):70, 2013. 14
 - [5] B. P. Abbott et al. A gravitational-wave standard siren measurement of the hubble constant. *Nature*, 551(7678):85–88, 2017. 14
 - [6] Rich Caruana, Steve Lawrence, and Lee Giles. Using machine learning to reconstruct the expansion history of the universe. In *Advances in Neural Information Processing Systems*, volume 18, pages 185–192, 2006. 14
 - [7] S. Spencer, T. Armstrong, J. Watson, S. Mangano, Y. Renier, and G. Cotter. Deep learning with photosensor timing information as a background rejection method for the cherenkov telescope array. *Astroparticle Physics*, 129:102579, 2021. 16
 - [8] Janis Fluri, Tomasz Kacprzak, Alexandre Refregier, Adam Amara, Aurelien Lucchi, and Thomas Hofmann. Cosmological constraints from noisy convergence maps through deep learning. *Physical Review D*, 98(12):123518, 2018. 16
 - [9] I. Gomez-Vargas, R. Medel-Esquivel, R. Garcia-Salcedo, and J. A. Vazquez. Neural network reconstructions for the hubble parameter, growth rate and distance modulus. *Eur. Phys. J. C*, 83:304, 2023. 17, 19
-

- [10] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tucillo, and J. L. Fischer. Improving galaxy morphologies for sdss with deep learning. *Monthly Notices of the Royal Astronomical Society*, 476(3):3661–3676, 2018. 19
- [11] Aizhan Akhmetzhanova, Siddharth Mishra-Sharma, and Cora Dvorkin. Data compression and inference in cosmology with self-supervised machine learning. *Journal of Cosmology and Astroparticle Physics*, 2023(02):053, 2023. 19
- [12] S. B. Maind and P. Wankar. Title of the paper. *IEEE International Journal on Recent and Innovation Trends in Computing and Communication*, 2(1):96–100, 2022. 20
- [13] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943. 20
- [14] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. 20
- [15] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. 20
- [16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 20
- [17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 21
- [18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 2672–2680, 2014. 22
- [19] S. H. Han, K. W. Kim, S. Kim, and Y. C. Youn. Artificial neural network: Understanding the basic concepts without mathematics. *Dement Neurocogn Disord*, 17(3):83–89, 2018. 22

- [20] L. Liu, X. Liu, J. Gao, H. Jiang, and J. Han. On the variance of the adaptive learning rate and beyond. *ICLR*, pages 1–13, 2021. 22
- [21] Ameya Jagtap and George Karniadakis. How important are activation functions in regression and classification? a survey, performance comparison, and future directions. 2022. 24
- [22] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 807–814, 2010. 25
- [23] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015. 26
- [24] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305, 2012. 27
- [25] François Chollet. *Deep Learning with Python, Second Edition*. MANNING, 2021. 28
- [26] Philipp Probst, Anne-Laure Boulesteix, and Bernd Bischl. Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, 20:1–32, 2019. 29
- [27] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, volume 24, pages 2546–2554, 2011. 29
- [28] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014. 30
- [29] Tonatiuh Matos. *¿De qué está hecho el Universo?* Fondo de Cultura Económica, 2004. 31
- [30] D. Comelli, M. Pietroni, and A. Riotto. Dark energy and dark matter. *Physics Letters B*, 571(3–4):115–120, 2003. 31

- [31] Andrew R. Liddle. *An Introduction to Modern Cosmology*. 1998. 31, 32
- [32] H. Karttunen, P. Kröger, H. Oja, M. Poutanen, and K. J. Donner. *Fundamental Astronomy, 6th ed.* Springer, 2016. 32
- [33] F. Zwicky. The redshift of extragalactic nebulae. *General Relativity and Gravitation*, 41(1):207–224, 2008. 32
- [34] G. J. Alner et al. Review of mathematics and physics of detection of wimps. *Reports on Progress in Physics*, 68(3):1521–1593, 2005. 32
- [35] S. van den Bergh. The hubble parameter. *Publications of the Astronomical Society of the Pacific*, 104:861, 1992. 33
- [36] European Space Agency. Planck: Resumen de la misión, 2013. 33
- [37] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, and et al. R. P. Kirshner. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998. 33
- [38] Nanoom Lee, Matteo Braglia, and Yacine Ali-Haïmoud. What it takes to solve the hubble tension through scale-dependent modifications of the primordial power spectrum. *arXiv preprint arXiv:2403.07574*, 2025. 34
- [39] P. J. E. Peebles and B. Ratra. The cosmological constant and dark energy. *Reviews of Modern Physics*, 75(2):559–606, 2003. 34
- [40] P. Bull, Y. Akrami, J. Adamek, T. Baker, E. Bellini, J. Beltrán Jiménez, E. Bentivegna, S. Camera, S. Clesse, J. H. Davis, E. di Dio, J. Enander, A. Heavens, L. Heisenberg, B. Hu, C. Llinares, R. Maartens, E. Mörtzell, S. Nadathur, and et al. H. A. Winther. Beyond λ cdm: Problems, solutions, and the road ahead. *Physics of the Dark Universe*, 12:56–99, 2016. 34
- [41] M. N. Castillo-Santos. Constricciones de energía oscura con enfoque en parametrizaciones fenomenológicas, 2023. 35, 36
- [42] M. N. Castillo-Santos, A. Hernández-Almada, M. A. García-Aspeitia, and J. Magaña. An exponential equation of state of dark energy in

- the light of 2018 cmb planck data. *Physics of the Dark Universe*, 40:101225, 2023. 35
- [43] Deng Wang and Xin-He Meng. Braneworld wormholes supported by astrophysical observations. *Front. Phys. (Beijing)*, 13:139801, 2018. 35
- [44] V. Motta, M. A. García-Aspeitia, A. Hernández-Almada, J. Magaña, and T. Verdugo. Taxonomy of dark energy models. *Universe*, 7(6):163–202, 2021. 35
- [45] A. Hernández-Almada. Cosmological test on viscous bulk models using hubble parameter measurements and type ia supernovae data. *Eur. Phys. J. C*, 79:751, 2019. 35
- [46] Planck Collaboration, N. Aghanim, et al. Planck 2018 results. vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020. 36
- [47] J. A. S. Lima, J. F. Jesus, R. C. Santos, and M. S. S. Gill. Is the transition redshift a new cosmological number? *arXiv Preprint*. 37
- [48] J. Liu, L. Qiao, B. Chang, and L. Xu. Revisiting cosmography via gaussian process. *Eur. Phys. J. C*, 83(374):1–10, 2023. 37
- [49] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, and D. Wilkinson. A 6of the epoch of cosmic re-acceleration. *Journal of Cosmology and Astroparticle Physics*, 2016. 38
- [50] Juan Magaña, Mario H. Amante, Miguel A. Garcia-Aspeitia, and V. Motta. The cardassian expansion revisited: constraints from updated hubble parameter measurements and type ia supernova data. *Monthly Notices of the Royal Astronomical Society*, 476:1036–1049, 2018. 38, 39