

### Objetivo de la red neuronal

El objetivo de esta red es poder predecir la calificación promedio de películas según su tiempo de duración, índice de aprobación y recaudaciones mundiales en cines, esto para poder determinar por medio de la calificación si a una película va a ser bien o mal criticada dependiendo de las variables mencionadas. Esta red utiliza regresión para poder calcular y predecir los efectos de las diferentes variables en la calificación promedio de las películas y poder dar como resultado una predicción con un nivel de precisión aceptable para poder determinar qué películas pueden ser exitosas o no con los críticos y el público general.

Se puede comentar que la red neuronal en este caso puede ser útil para realizar este tipo de predicciones debido a que puede realizar manejo adecuado de relaciones no lineales entre las variables de entrada y la calificación promedio de las películas (Variable objetivo), lo que puede resultar en una captación de relaciones más complejas entre estas variables para predecir los resultados con una mayor precisión. Adicionalmente tiene flexibilidad para el aprendizaje y la escalabilidad del modelo por si se quisieran realizar predicciones con nuevos variables a tomar en cuenta o utilizar conjuntos de datos con más información para aumentar la precisión de las predicciones.

### Subconjunto de variables a utilizar

Movie\_average: Tipo Número (Variable Objetivo)

Esta variable es el objetivo de predicción de la red neuronal planteada con regresión, que por medio de los valores encontrados en las demás variables que se listaran posteriormente se intentará realizar un cálculo predictivo con la red con el objetivo de tener una predicción de la calificación de las películas.

Runtime\_minutes: Tipo Entero

Esta variable se escogió debido a que dependiendo de si una película es muy larga o muy corta se puede encontrar relaciones con lo que las personas piensan de las películas debido a aburrimiento por ser muy larga o si es muy corta puede dejar a personas con ganas de más.

Approval\_index: Tipo Número

Esta variable se escogió debido a que está relacionada a la movie\_avg y se puede utilizar para predecir el promedio de la película de forma más sencilla.

Worldwide\_gross: Tipo Entero

Esta variable se escogió debido a que nos puede ayudar a predecir el movie\_avg viendo si hay relación con la cantidad de personas que vieron la película con su punteo.

## Resultados

- Red 1

Para la primera red neuronal se quiso construir algo simple para observar cómo se comporta el dataset y las predicciones ante esto, esta red consta de principalmente 2 capas, una capa oculta con 32 neuronas y 1 neurona en la capa de salida, la capa es una capa oculta densa conectada con las tres entradas, esta capa utiliza la función de activación 'relu', la cual es una función de activación no lineal que se aplica elemento a elemento, que permite aprender relaciones y patrones más complejos en los datos. Por simplicidad esta red se decidió realizar sin ninguna técnica de regularización.

Debido a la simplicidad del modelo/red empleada el entrenamiento del mismo tomo un tiempo relativamente corto de 5.8 segundos para 50 épocas, tras realizar una evaluación con los datos de prueba se encontró un MSE de 0.26, lo cual considerando el tipo de datos que se quiere predecir se puede considerar que el modelo es bastante preciso.

```
1/69 [.....] - ETA: 1s - loss: 0.1948  
69/69 [=====] - 0s 1ms/step - loss: 0.2609  
MSE en el conjunto de prueba Modelo 1: 0.260897159576416
```

Fig. 1. MSE del modelo 1

En la figura 2 se puede observar como el modelo fue evolucionando a través de las épocas utilizando el conjunto de entrenamiento para reducir su MSE con el paso de cada época. Esta figura es en donde principalmente se puede llegar a apreciar mejor el propio funcionamiento del entrenamiento del modelo y según la tendencia observada del modelo se podría decir que da predicciones dentro de un margen de error aceptable.

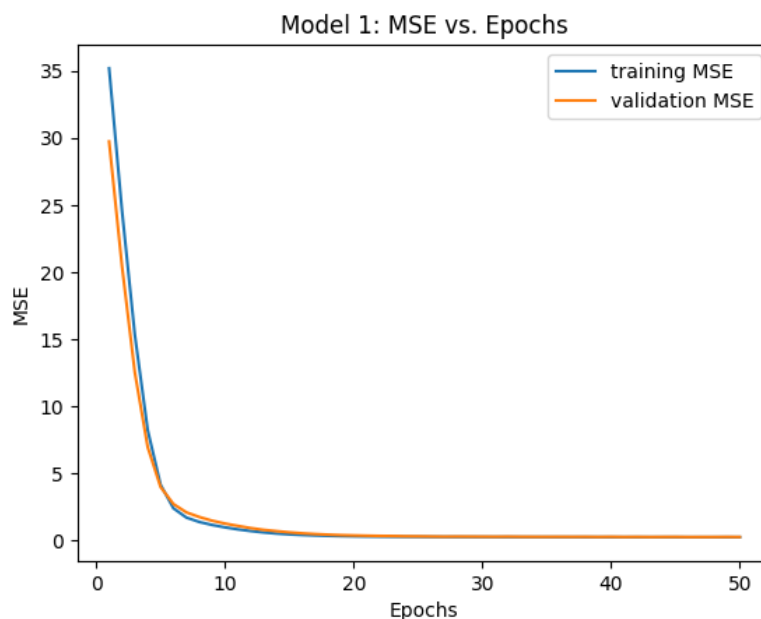


Fig. 2. Modelo 1: MSE vs. Épocas

- Red 2

Para esta red neuronal se eligió realizar un modelo un poco más complejo que el anterior, contando con 2 capas ocultas, contando con 32 y 12 neuronas respectivamente y adicional a esto se cuenta con 1 capa de salida con una neurona, en esta ocasión se eligió utilizar la función de activación 'relu' al igual que el modelo anterior, introduciendo la no linealidad a la red. El modelo cuenta con 2 capas de normalización, una tras la capa oculta 1 la cual es una Batch Normalization, la cual se encarga de normalizar las activaciones de la capa oculta 1 al ajustar su media y varianza, lo que puede ayudar a mejorar la estabilidad y el entrenamiento de la red, esto ayuda a acelerar la convergencia y puede reducir el problema de gradientes que se desvanecen. También, se cuenta con una segunda capa de batch normalization tras la capa oculta 2.

Se obtuvo un tiempo total de entrenamiento de 8.2 segundos, además, se consiguió un MSE de 0.24, lo cual se podría considerar como marginalmente un mejor resultado que el anterior, esta disminución en el MSE probablemente se deba a la normalización y la mayor complejidad de la red a comparación de la red 1.

```
MSE en el conjunto de prueba Modelo 1: 0.260897159576416  
69/69 [=====] - 0s 912us/step - loss: 0.2472  
MSE en el conjunto de prueba Modelo 2: 0.2471781075000763
```

Fig. 3. MSE del modelo 2

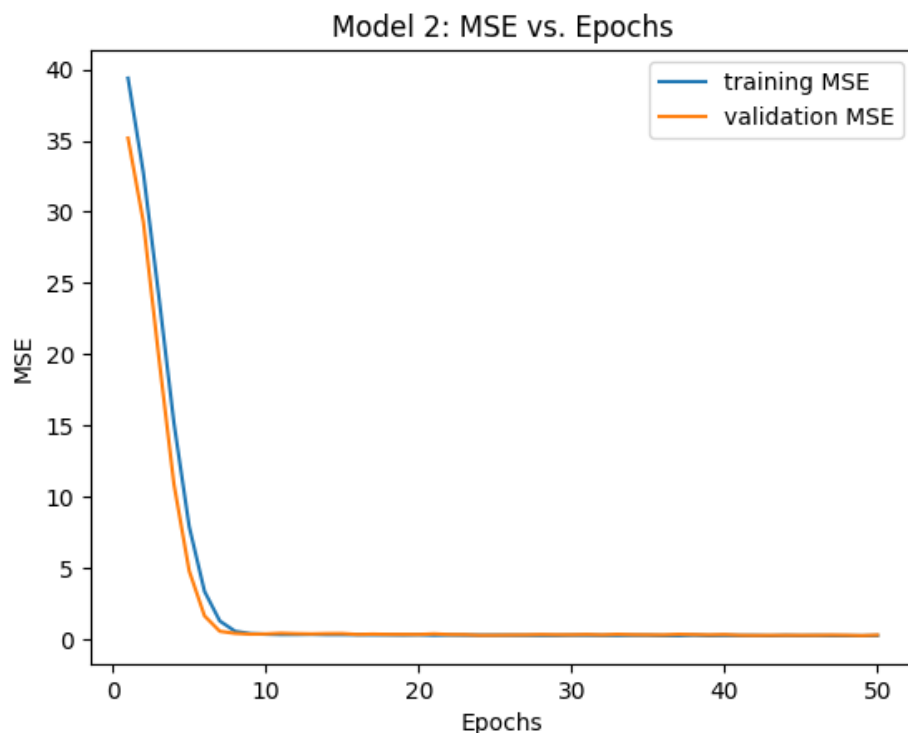


Fig. 4. Modelo 2: MSE vs Épocas

- Red 3

Para esta red se decidió hacer un modelo con una complejidad considerablemente mayor a los anteriores dos, para observar si se disminuía aún más el MSE, se utilizaron 6 capas ocultas, con una capa de salida. Las capas ocultas cuentan cada una con 32, 64, 128, 64 y 32 neuronas respectivamente. Todas las capas ocultas utilizan la función de activación ‘tanh’, que mapea los valores de entrada en el rango (-1,1) y de manera similar a ‘relu’ introduce no linealidad en la red, pero esta vez introduciendo la posibilidad de obtener salidas en el rango negativo y positivo. Adicionalmente, similar a la red 2, se cuenta con capas de Batch Normalization entre cada una de las capas ocultas.

Aun cuando se aumentó mucho la complejidad y se esperaba un MSE que disminuyera bastante a comparación de las demás redes se pudo observar que aumentó de manera considerable, llegando a un MSE de 0.34, aunque este no sea un MSE el cual sea pésimo, es menos viable que las otras dos debido a su MSE y que además su consumo de recursos (tiempo de entrenamiento) es mayor a las dos redes anteriores; 12.2 segundos debido a su complejidad.

```
MSE en el conjunto de prueba Modelo 2: 0.2471781075000763  
69/69 [=====] - 0s 1ms/step - loss: 0.3464  
MSE en el conjunto de prueba Modelo 3: 0.34635376930236816
```

Fig. 5. MSE del modelo 3

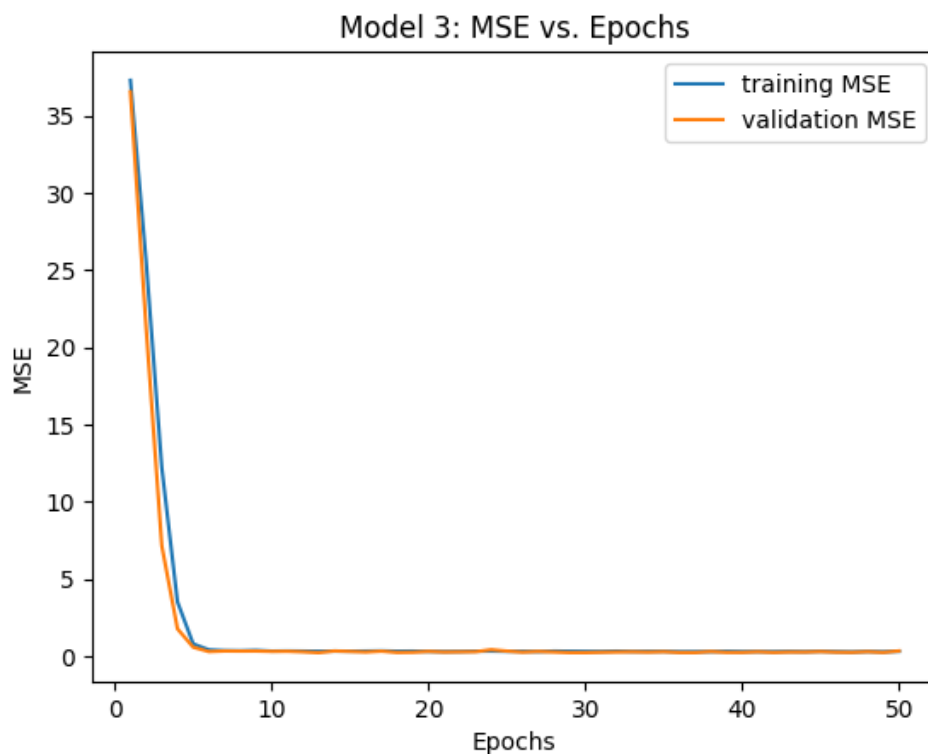


Fig. 6. Modelo 3: MSE vs. Epocas

### Discusión

Tras el análisis y comparación de las tres redes neuronales, cada una conteniendo su propio nivel de complejidad para la tarea de predicción de un conjunto de datos. La red 1, aun siendo la más simple, consistiendo de una capa oculta con 32 neuronas, logró conseguir un MSE de 0.26 y requirió un tiempo relativamente corto para su entrenamiento.

La red 2, con dos capas ocultas de 32 y 12 neuronas incorporó técnicas de normalización (batched normalization) y logró mejorar ligeramente el MSE, este siendo 0.24, lo que probablemente significa que la normalización tuvo un impacto positivo en la precisión de la red, se observó un tiempo relativamente mayor al de la red 1 debido a su mayor complejidad.

La red 3, a pesar de ser la más compleja, con seis capas ocultas y un mayor número de neuronas a comparación de las otras dos redes, no logró mejorar el MSE y al contrario, empeoro hacia un valor de 0.34, siendo el más alto de las 3 redes propuestas, además que se obtuvo un tiempo de entrenamiento más largo que los otros dos con 12.2 segundos.

En cuanto a la elección de la función de activación 'relu' en las primeras dos redes parece haber funcionado bastante bien, por lo que se puede inferir que esta es la decisión correcta en este aspecto. Por otro lado, la función de activación 'tanh' no parece haber beneficiado el rendimiento de la red 3, a comparación de 'relu' para las redes 1 y 2.

La introducción de la batch normalization, en la red 2 parece haber mostrado cierta mejora en el MSE a comparación de la red 1, por lo que esto respalda el concepto de que la normalización puede llegar a mejorar la estabilidad del modelo.

En general, los resultados nos dicen que aumentar la complejidad de golpe del modelo no siempre conduce a una mejora significativa en precision, mas bien es importante que se encuentre un equilibrio entre en modelo/red y los datos que se están utilizando para obtener que nos de resultados lo más óptimos posibles y evitar llegar a resultados subóptimos, un modelo underfitted o un modelo overfitted.

En un futuro podría ser interesante explorar más de las posibilidades para cada uno de los modelos con el objetivo de encontrar un modelo que nos provea con la mejor relación posible de rendimiento/resultados y que tenga una mayor robustez en cuanto a predicciones con datos fuera del entrenamiento.

### La mejor red neuronal

La mejor elección entre las redes neuronales propuestas en el laboratorio es la red 1, debido a la simplicidad, sus resultados y su tiempo de entrenamiento, ya que el tener un MSE bajo como además de un tiempo de entrenamiento relativamente bajo lo hace propenso expandirse a ser un modelo bastante viable cuando se entrena con un dataset mucho más grande y que podría llegar hasta a mejorar su precisión. Por lo que por sus características de ser de bajo costo para el entrenamiento y tener una precisión bastante buena entre las redes

analizadas podría ser la elección como mejor (o por lo menos la más optima) red neuronal para el dataset estudiado en conjunto con los objetivos planteados para el laboratorio.