



Cesar Alejandro Cruz Salas

A00825747

Inteligencia Artificial Avanzada para la ciencia de datos

Modulo Aprendizaje Máquina

“Portafolio Análisis”

Profesor:

Jesús Adrián Rodríguez Rocha

Resumen

En este informe, se presenta el desarrollo de una Red Neuronal Recurrente (RNN) Long Short Term Memory (LSTM) diseñada para la predicción de los valores de las acciones de "Apple". El análisis se inicia con la implementación de un modelo inicialmente sencillo. Luego, se abordan y solucionan los desafíos relacionados con el sesgo (bias), la varianza (variance), el subajuste (underfitting) y el sobreajuste (overfitting), lo que resulta en una mejora significativa del modelo.

1.-Introducción

En este informe, se explora el desarrollo de una Red Neuronal Recurrente (RNN), en particular una Long Short Term Memory (LSTM), con el propósito de llevar a cabo la predicción de los valores de las acciones de la empresa "Apple". Esta elección de arquitectura se basa en el conocimiento adquirido durante la formación, donde se enfatizó la eficacia de las LSTM para abordar problemas vinculados a series temporales.

Inicialmente, se configura un modelo que consta de dos capas LSTM, cada una compuesta por 10 neuronas, seguidas de tres capas densas, cada una integrada por 5 neuronas. Dado que la tarea consiste en predecir un único valor de salida, se utiliza una sola neurona en la capa de salida.

Para mejorar la precisión del modelo, se aplican estrategias clave. Se introduce la técnica de Dropout, destinada a reducir el sobreajuste desconectando de manera aleatoria algunas conexiones neuronales durante el entrenamiento. También se incorpora Batch Normalization, con el propósito de estabilizar el proceso de entrenamiento y acelerar la convergencia del modelo. Adicionalmente, se emplean callbacks, como Early Stopping y la reducción de la tasa de aprendizaje, con el fin de evitar el sobreajuste y optimizar el rendimiento del modelo a lo largo de múltiples épocas.

La métrica del Error Cuadrático Medio de Raíz (RMSE) se emplea para evaluar la precisión del modelo en la predicción de los precios de las acciones de Apple. Estas mejoras y enfoques se aplican con el propósito de obtener resultados más precisos y confiables en la predicción de series temporales.

2.-Separación y evaluación del modelo con un conjunto de prueba y un conjunto de validación (Train/Test/Validation):

En el código proporcionado, se observa que se separa el conjunto de datos en un conjunto de entrenamiento (train_data) y un conjunto de prueba (test_data) utilizando el 80% de los datos para entrenar y el 20% para test. La red neuronal implementada utiliza el 15% de los datos de entrenamiento para validar el modelo y utiliza como métrica el Error Cuadrático Medio de Raíz (RMSE).

3.-Diagnóstico y explicación del grado de bias o sesgo:

El nivel de sesgo en el modelo es intrínseco a la arquitectura seleccionada. En particular, el "model1" inicial presenta una estructura relativamente sencilla, compuesta por dos capas LSTM de 10 neuronas y tres capas densas, cada una con 5 neuronas. Al analizar la Figura 1, se observa que este modelo, en comparación con los otros implementados, especialmente en contraste con el tercero, que muestra un comportamiento más caótico, exhibe una tendencia. Concretamente, se aprecia que el error durante el entrenamiento tiende a ser muy reducido, prácticamente

alcanzando un valor cercano a cero. Sin embargo, al examinar el error de validación, se constata que es notablemente elevado en comparación con el error de entrenamiento. Este fenómeno sugiere que, aunque el modelo es más simple en su estructura, podría estar sufriendo de sobreajuste o "overfitting".

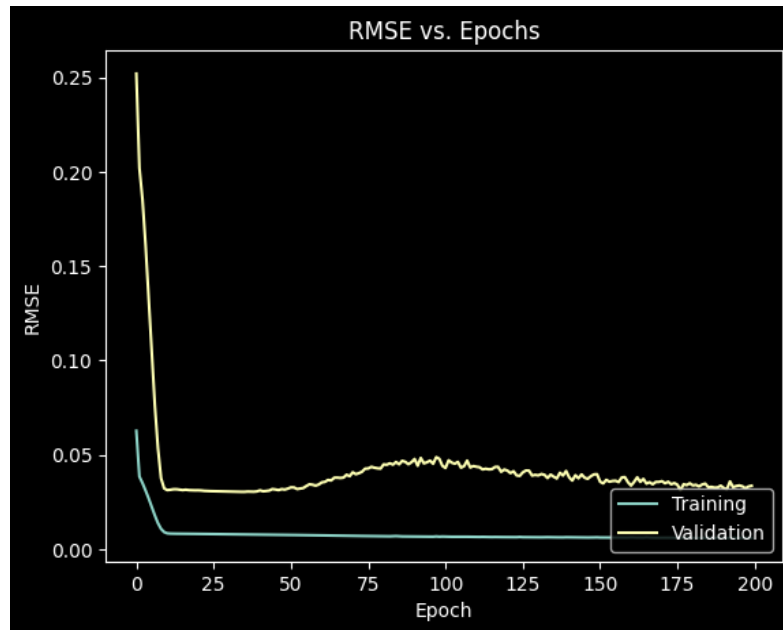


Figure 1.- RMSE vs Epochs- Model 1

4.-Diagnóstico y explicación del grado de varianza:

En el "model 2", se implementan técnicas de regularización, como Batch Normalization y Dropout, con el objetivo de abordar el problema de sobreajuste en el modelo. Al observar la figura 2, se puede notar que aunque el error en el conjunto de entrenamiento aumenta en comparación con el "model 1", el aumento en el error de validación es significativamente menor en relación al primer modelo. Sin embargo, es importante señalar que aún persiste una brecha entre ambas curvas de error, y el error de validación sigue siendo más alto en ambos casos.

Estos resultados sugieren que, aunque se ha logrado reducir el sobreajuste en cierta medida con las técnicas de regularización implementadas en el "model 2", todavía no se ha alcanzado una generalización óptima para datos nuevos. Esto indica que la varianza en el modelo sigue siendo alta, lo que implica que el modelo no generaliza bien para datos que no ha visto previamente.

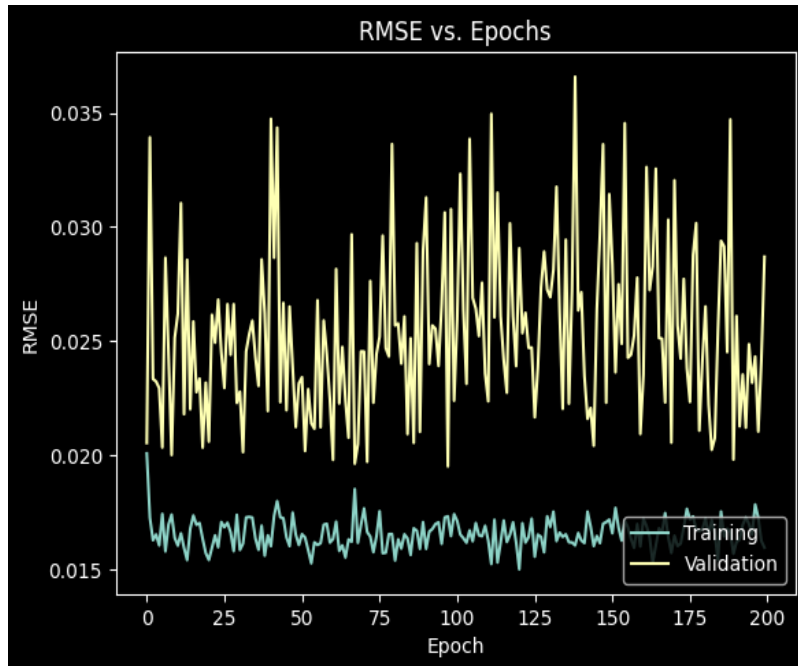


Figure 2.- RMSE vs Epochs- Model 2

5.-Diagnóstico y explicación del nivel de ajuste del modelo (Underfitt-Overfitt)

Al analizar exhaustivamente los modelos implementados, se tomaron medidas significativas para abordar el problema de overfitting. Se incorporaron estrategias clave, como la implementación de Early Stopping después de 30 épocas y la reducción de la tasa de aprendizaje después de 20 épocas, lo que se tradujo en una marcada disminución del error, como se ilustra en la figura 3. Sin embargo, a pesar de estas mejoras, se observa que el modelo aún presenta indicios de overfitting.

Para contrarrestar este fenómeno, se reconoce la necesidad de reducir la complejidad del modelo. Cabe mencionar que, en la búsqueda de soluciones, se realizaron pruebas experimentales con un solo epoch y un batch size de 1. Estos ajustes se llevaron a cabo con el propósito de acelerar mejorar el entrenamiento, ya que, en la configuración mencionada se tienen mejores resultados, pero tiempos más extensos de ejecución y en este reporte se buscaba ver como se comportaba el modelo a través de los epochs.

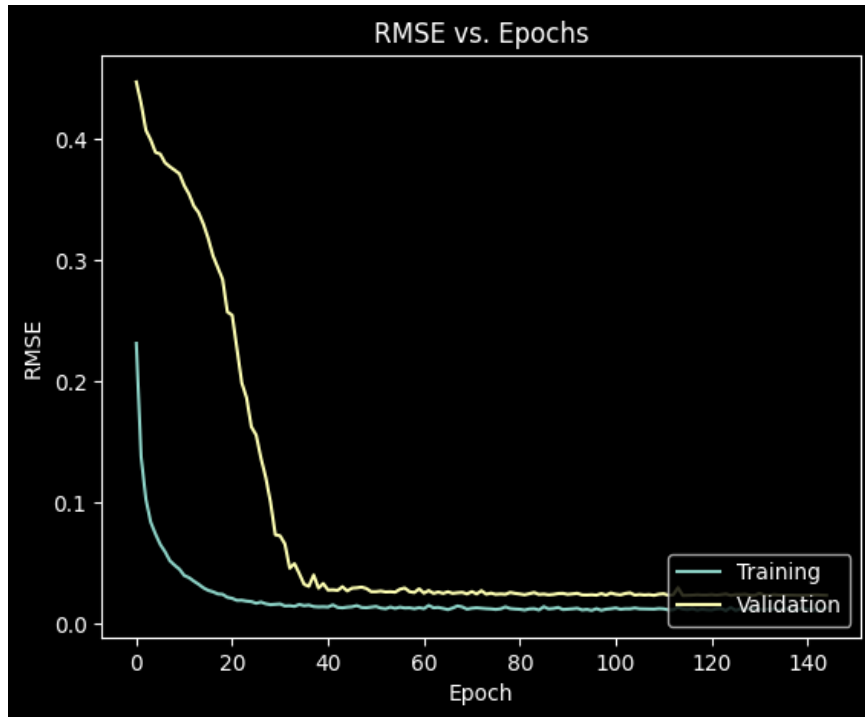


Figure 3.- RMSE vs Epochs- Model 3

6.Conclusion.

A lo largo de este proyecto de predicción de las acciones de Apple, se aprendió a identificar y manejar de manera efectiva el sesgo, el overfitting y la varianza en los modelos de redes neuronales recurrentes (RNN). Se descubrió que un modelo inicialmente sencillo, como el "model 1", puede mostrar un alto sesgo y subajuste, ya que tiende a simplificar en exceso la complejidad de los datos, lo que se refleja en un error de validación significativamente alto en comparación con el entrenamiento.

El overfitting, por otro lado, se convirtió en un desafío central a medida que aumentaba la complejidad del modelo. Aprendí que estrategias como el Dropout y Batch Normalization son fundamentales para mitigar este problema, pero aún así, es esencial encontrar un equilibrio adecuado para evitar un ajuste excesivo a los datos de entrenamiento. La varianza también se destacó como un factor crítico en la generalización de los modelos y comprendí que, incluso con técnicas de regularización, reducir la varianza sin sacrificar la calidad de las predicciones es un desafío constante en este tipo de tareas de predicción de series temporales. En resumen, este proyecto me ha enseñado la importancia de equilibrar sesgo y varianza en modelos de RNN para lograr predicciones más precisas y confiables en el mercado de acciones.

7.Referencias

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.

8.Anexo

Google drive: https://drive.google.com/drive/folders/1Y1v7Scb-gWuNms_oy2BTX_uEZMjQQ40U?usp=sharing