Apuntes De Clase Semana 5
Inteligencia Artificial

Perez Picado Esteban, *2021046572*

***Abstract*—This document presents key concepts discussed in week 5 of the Artificial Intelligence course. It covers recent advancements in AI, including Meta's research on transformer models without normalization, developments in robotics by Boston Dynamics, and a study on AI agents by Oracle. Additionally, it provides an overview of essential readings on fundamental algorithms and learning models. A detailed description of the first assignment is included, focusing on linear regression with gradient descent and feature selection using Kaggle datasets. The document also explores key concepts in bias-variance tradeoff, data splitting techniques, and logistic regression, emphasizing the sigmoid function and optimization methods.**

I. NOTICIAS

A. Paper de Meta

El reciente artículo de Meta, titulado "Transformer without Normalization", aborda una optimización en la aplicación de transformadores en los modelos de lenguaje de gran tamaño (LLMs). Tradicionalmente, la salida de cada capa se normaliza utilizando técnicas ampliamente reconocidas por su efectividad. Sin embargo, este proceso conlleva un alto costo computacional. La investigación revela que la salida de cada capa sigue una forma similar a una "S", lo que permite proponer un método alternativo de normalización basado en la función hiperbólica. Esta técnica también contribuye a la reducción del costo de inferencia y entrenamiento.

B. Video de Manus

Se destaca un video de Manus, un divulgador de inteligencia artificial en español recomendado por OpenCCV. En este contenido, se presenta una herramienta que permite clasificar papers por categoría utilizando técnicas de Web Scraping. Mediante esta herramienta, Manus analiza un artículo y categoriza automáticamente cada uno de los papers referenciados en él, proporcionando una visión estructurada y organizada de la información.

C. Evolución de la Inteligencia Artificial

Actualmente, la inteligencia artificial experimenta un crecimiento exponencial, duplicando sus capacidades aproximadamente cada siete meses. Esta velocidad de avance es tan acelerada que muchas veces no se dispone de tiempo suficiente para comprender en profundidad los nuevos desarrollos.

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

D. Nuevo Robot de Boston Dynamics

Boston Dynamics ha presentado su nuevo robot, denominado Atlas, equipado con un traje de movimiento que le permite aprender de los movimientos humanos mediante inteligencia artificial. Gracias a esta capacidad, Atlas puede imitar con gran precisión y fluidez los gestos y movimientos humanos, marcando un avance significativo en el desarrollo de la robótica autónoma.

E. Estudio de Agentes de Oracle

Oracle ha publicado un estudio sobre la producción de agentes de inteligencia artificial. Además, se está considerando la migración de PyTorch a la Linux Foundation. No obstante, Yann LeCun ha manifestado su oposición a esta medida, argumentando que, en caso de concretarse, Meta retiraría su soporte al proyecto, lo que podría afectar su desarrollo y evolución en la comunidad de código abierto.

II. LECTURAS

En la plataforma Tec Digital, ya está disponible el capítulo 3 de "Fundamental Algorithms", el cual aborda conceptos esenciales como algoritmos fundamentales, regresión lineal y regresión logística, la función sigmoide y la verosimilitud, entre otros temas relevantes. Este capítulo también cubre algoritmos que, aunque no necesariamente se verán en clase, son importantes para comprender el desarrollo de la inteligencia artificial.

Asimismo, el capítulo 4 de "Anatomy of a Learning Algorithm" profundiza en aspectos como gradientes de pendiente, derivadas parciales en función de los parámetros, épocas y la tasa de aprendizaje. Estos conceptos están directamente relacionados con los contenidos vistos en la clase pasada, proporcionando un complemento útil para reforzar el aprendizaje.

III. TAREA 1

La primera tarea debe entregarse el 27 de marzo de 2025 y tiene un valor de 3.33. Posteriormente, quedarán solo dos tareas más con el mismo valor. Esta tarea consiste en aplicar regresión lineal con descenso de gradiente y practicar la selección de características (features). Para ello, se utilizará Kaggle como fuente de datos, descargando un conjunto de datos sobre valores de viviendas. En el enlace proporcionado se encontrará información sobre el conjunto de datos, la explicación de cada característica, como el número de habitaciones o baños, y lo más importante: la etiqueta, que es la variable dependiente con la cual se realizarán las predicciones.

El conjunto de datos también contiene notebooks con predicciones previas, los cuales pueden ser utilizados como

referencia. Sin embargo, el objetivo principal es descubrir nuevas técnicas de análisis de datos y emplear todas las herramientas posibles para mejorar el desempeño del modelo.

El procesamiento de datos, como la conversión de unidades (por ejemplo, de pies a metros cuadrados), es una opción válida, pero debe justificarse adecuadamente. Toda la explicación y análisis se realizarán en un notebook, sin necesidad de documentación en LaTeX.

En cuanto a las características ordinales, como la calidad de los vecindarios, se evaluará su impacto en las regresiones para decidir si conservarlas, eliminarlas o transformarlas. Para las características numéricas, se aplicará análisis de correlación para seleccionar únicamente aquellas que aporten información relevante y evitar problemas de multicolinealidad. La meta es reducir la cantidad de características irrelevantes y optimizar la predicción.

Para evaluar el sesgo y la varianza del modelo, se dividirá el conjunto de datos y se analizará el modelo con el error cuadrático medio (MSE). Se realizará un seguimiento del comportamiento del entrenamiento en cada época y se generarán curvas de aprendizaje para detectar posibles problemas de sobreajuste (overfitting) o subajuste (underfitting). En caso de detectar estos problemas, se incluirán celdas adicionales en el notebook con estrategias para corregirlos.

IV. SESGO Y VARIANZA

A. Validation Set

Para controlar el sesgo y la varianza, es importante evitar el sobreajuste (overfitting), que ocurre cuando el modelo aprende de memoria los datos del conjunto de entrenamiento y no generaliza adecuadamente a nuevos casos.

El conjunto de prueba (testing set) se utiliza para evaluar el desempeño del modelo mediante métricas como la precisión (accuracy) o la función de pérdida (loss). Para evitar que el modelo presente sobreajuste después de un largo entrenamiento, se introduce el conjunto de validación (validation set). Este conjunto permite monitorear el progreso del modelo durante el entrenamiento y detectar problemas de sobreajuste de manera temprana. Además, es fundamental para ajustar los hiperparámetros y mejorar la capacidad de generalización del modelo.

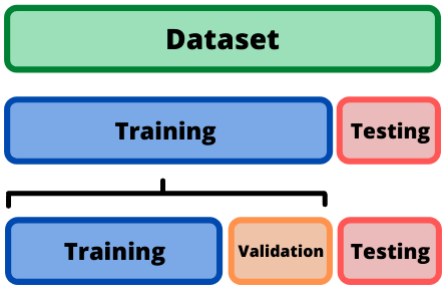

Fig. 1. Flujo de división del dataset.

B. Técnicas de División de Datos

Existen tres técnicas principales para dividir los datos:

- **Random Sampling**: Divide aleatoriamente el conjunto de datos. Es útil cuando los datos están equilibrados, pero puede generar sesgos si las clases están desbalanceadas.
- **Stratified Sampling**: Asegura que todas las clases estén representadas en la misma proporción en cada conjunto de datos.
- **K-Fold Cross Validation**: Divide el conjunto de datos en *k* partes, utilizando *k-1* partes para el entrenamiento y la restante para la validación. El proceso se repite rotando las partes utilizadas para cada fase.

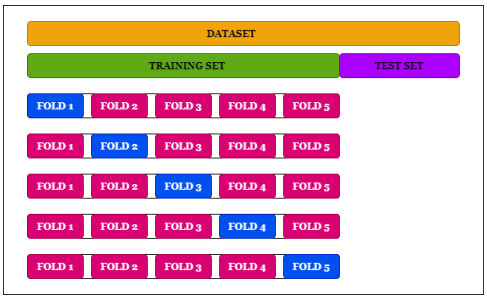

Fig. 2. Técnica y proceso del K-Fold Cross Validation.

C. Escenarios para los Resultados del Modelo

Los resultados del modelo pueden presentar distintos escenarios:

- **Escenario ideal**: La función de pérdida disminuye tanto en el entrenamiento como en la validación, indicando un buen ajuste del modelo.

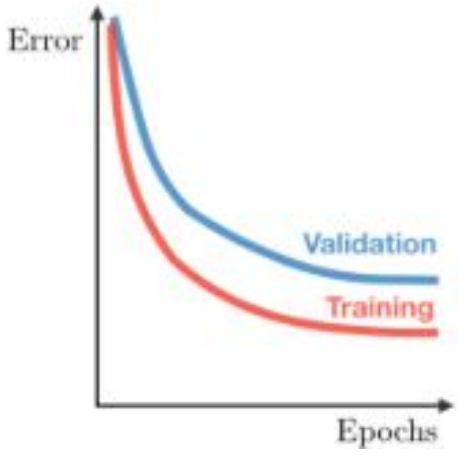

Fig. 3. Escenario ideal entre validación y entrenamiento.

- **Sobreajuste (overfitting)**: El error en el entrenamiento es bajo, pero el error en la validación es alto, lo que significa que el modelo no generaliza bien.

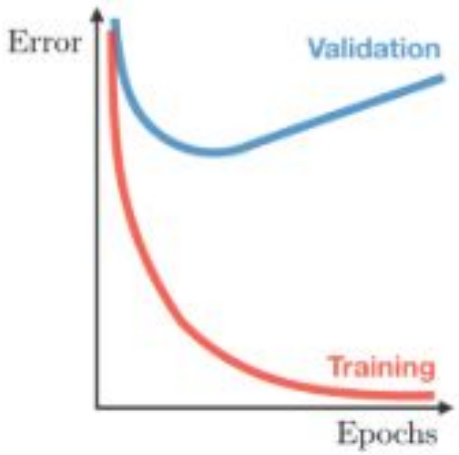

Fig. 4. Escenario con overfiting.

- **Subajuste (underfitting)**: Tanto el error en entrenamiento como en validación son altos, lo que indica que el modelo es demasiado simple y no aprende adecuadamente.

Fig. 5. Escenario con overfiting.

D. Balance entre Sesgo y Varianza

El balance entre sesgo y varianza es crucial. Un modelo más complejo tiende a tener alta varianza, mientras que un modelo más simple puede tener alto sesgo. Para evitar un sesgo elevado, se puede optar por un modelo más complejo o seleccionar características más representativas. En casos de alta varianza, se recomienda simplificar el modelo, reducir la dimensionalidad, obtener más datos o aplicar técnicas de regularización.

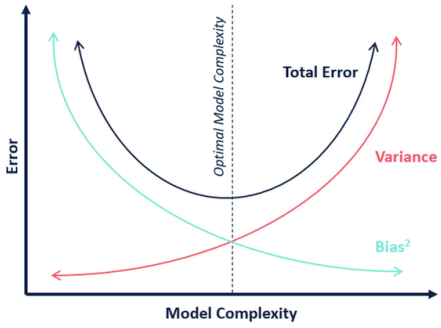

Fig. 6. Comparación de la complejidad del modelo.

V. REGRESIÓN LOGÍSTICA

Con respecto a la regresión logística, debemos tener en cuenta que esta y la regresión lineal son muy diferentes. La regresión lineal genera un modelo de valores continuos, mientras que la regresión logística se emplea para la clasificación binaria. Un ejemplo de esto es que, dado el tamaño de una calabaza, podríamos usar la regresión lineal para estimar su precio, mientras que con la regresión logística podríamos predecir si es de color naranja o no, proporcionando la probabilidad de que ocurra dicho evento.

Dado un conjunto de datos, se marcan las observaciones con 0 o 1 según las características presentes. La distribución sigue una distribución Bernoulli, donde k puede tomar valores de 0 o 1, y p representa la probabilidad de que ocurra el evento.

DISTRIBUCIÓN DE BERNOULLI

La función de masa de probabilidad de una variable aleatoria con distribución de Bernoulli es:

$$P(X = k) = p^k \cdot (1 - p)^{1-k}$$

donde $k \in \{0, 1\}$ y p es la probabilidad de éxito.

A. Función Sigmoidal

La función sigmoidal o función logística estándar tiene un comportamiento no lineal con un rango de valores en $[0, 1]$, aceptando cualquier número real como entrada. Esta función permite transformar una combinación lineal de características en una probabilidad.

FUNCIÓN SIGMOIDE

La función sigmoide se define como:

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

Esta función transforma cualquier número real en un valor en el rango $(0, 1)$.

Si tomamos la idea de la regresión lineal y aplicamos una función sigmoide sobre su resultado, obtenemos un modelo de clasificación. Esto introduce no linealidad sin requerir una computación costosa, lo que permite resolver problemas de mayor complejidad de manera eficiente.

LINEALIDAD

Una función lineal se expresa como:

$$f_{w,b}(x) = wx + b$$

donde w es el peso y b es el sesgo.

Al aplicar la función sigmoide a esta transformación lineal, obtenemos:

$$\sigma(f_{w,b}(x)) = \frac{1}{1 + e^{-f_{w,b}(x)}}$$

Para realizar la clasificación, establecemos un umbral: si la función sigmoidal es mayor o igual a 0.5, la clasificación es 1; en caso contrario, es 0. Este umbral puede ajustarse según el problema y sus necesidades específicas.

FUNCIÓN SIGMOIDE

$$f_{w,b}(x) = \frac{1}{1 + e^{-(wx+b)}}$$

B. Ejemplo con Neuronas

En una red neuronal, los valores de entrada (features) ponderados por sus respectivos pesos son sumados junto con un sesgo. Luego, esta suma pasa por una función sigmoide, la cual introduce la no linealidad. Finalmente, la salida aproximada representa la predicción del modelo.

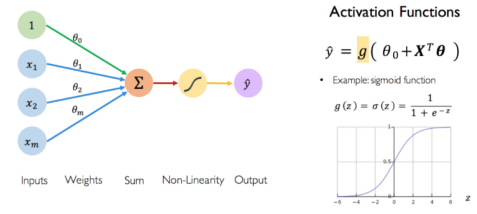

Fig. 7. Ejemplo de una red neuronal.

C. Optimización de Pesos

Para ajustar los pesos w y el sesgo b en la regresión logística, primero se define una función de pérdida. Posteriormente, se calculan las derivadas parciales respecto a los parámetros y se actualizan mediante algoritmos de optimización como el descenso de gradiente. La derivada de la función sigmoide se expresa como $\sigma(x) \cdot (1 - \sigma(x))$.

DERIVADA DE LA FUNCIÓN SIGMOIDE

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

Regla del cociente: si f(x) = u(x) / v(x), entonces su derivada es:

$$\left(\frac{u(x)}{v(x)}\right)' = \frac{u'(x)v(x) - u(x)v'(x)}{v(x)^2}$$

$$\sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

$$\sigma'(x) = \frac{e^{-x} + 1 - 1}{(1 + e^{-x})^2}$$

$$\sigma'(x) = \frac{e^{-x} + 1}{(1 + e^{-x})^2} - \frac{1}{(1 + e^{-x})^2}$$

$$\sigma'(x) = \frac{1}{(1 + e^{-x})}\left(1 - \frac{1}{(1 + e^{-x})}\right)$$

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$$

D. Función de Pérdida

En lugar de minimizar el error cuadrático medio (MSE), la regresión logística utiliza la función de verosimilitud, ya que maneja probabilidades. La verosimilitud mide la probabilidad de observar los datos dados ciertos parámetros θ.

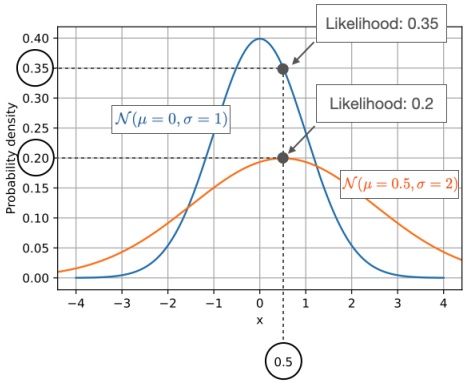

Fig. 8. Afectación de la varianza de los datos.

PROBABILIDAD DE OBSERVAR TODOS LOS DATOS

$$L(\theta|X) = P(x_1, x_2, \ldots, x_n|\theta) = P(x_1|\theta) \cdot P(x_2|\theta) \cdots P(x_n|\theta)$$

Si los datos se distribuyen normalmente y nuestros ejemplos están desplazados con respecto a la distribución, la probabilidad de observarlos será baja. El objetivo es maximizar la verosimilitud para optimizar los parámetros w y b, asegurando que el modelo explique los datos de la mejor manera posible.

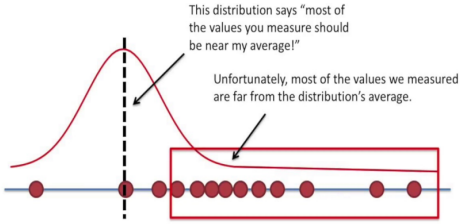

Fig. 9. Desviación de una función normal a la izquierda.