Apuntes Semana 7, Martes 19 de marzo de 2024

1st Mauricio Agüero Márquez
Escuela de Computación
Instituto Tecnológico de Costa Rica
Cartago, Costa Rica
mam.121002@estudiantec.cr

I. Quiz #3

A. Describa la técnica y fórmula de Normalización y Estandarización. (35pts)

Normalización: Tomar el rango de valores de un dato y ajustarlo a un intervalo estándar. Por lo general el intervalo que se busca obtener es [0, 1] o [-1, 1].

Estandarización: Transformar los valores para que la distribución de los datos tenga una media de 0 y una desviación estándar de 1, es decir, que los datos sigan una distribución normal.

B. Describa qué es overfitting y underfitting. (25pts)

Overfitting: Ocurre cuando un modelo es demasiado simple para capturar la estructura subyacente de los datos de entrenamiento. Esto se manifiesta en una baja capacidad del modelo para predecir con precisión incluso los datos de entrenamiento. Suele ocurrir cuando el modelo es demasiado básico o porque las características utilizadas para el entrenamiento no son lo suficientemente informativas.

Underfitting: Se produce cuando un modelo es demasiado complejo y se ajusta demasiado a los datos de entrenamiento; aunque aprende bien de los datos de entrenamiento, generalmente clasifica mal en los datos de prueba. Las causas principales del overfitting incluyen modelos demasiado complejos o un número insuficiente de ejemplos de entrenamiento.

C. Nombre y explique el proceso de convertir features continuos a múltiples features binarios. (20pts)

Binning: Este proceso se basa en dividir el rango de valores de un feature continuo en intervalos, también conocidos como bins. Por ejemplo, si el feature que se quiere dividir es la edad, el rango de 0-5 años se podría definir como el primer bin, 6-10 años como el segundo bin, 11-15 años como el tercero, y así sucesivamente. Por cada bin, se crea una nueva característica binaria que representa si el valor del feature original corresponde al intervalo del bin (1) o no (0).

D. Defina qué es un epoch. (20pts)

Epoch: Es una iteración completa a través de todo el training set. Durante un epoch, el algoritmo procesa todos los ejemplos del conjunto de entrenamiento y actualiza sus parámetros en base a lo que ha aprendido.

II. Feature Scaling

Este concepto se refiere a la modificación del rango de valores de un feature para manipular mejor la información del dataset. Este proceso es necesario en algoritmos que involucran un cálculo de gradiente ya que ayudan a que converga más rápido porque los datos están en una escala similar. También ayuda a mejorar la precisión al asegurarse de que los datos tengan una baja varianza.

No es necesario hacer un escalado cuando el algoritmo se basa en distancias como con el k-means, decision trees o support vector machine (SVM).

A. Normalización

En la normalización se manipula un rango de valores de un dato y para que la información esté en un rango específico. Usualmente ese rango es [0, 1] o [-1, 1]. En Fig. 1 se visuaiza la diferencia entre un feature antes y después de ser normalizado.

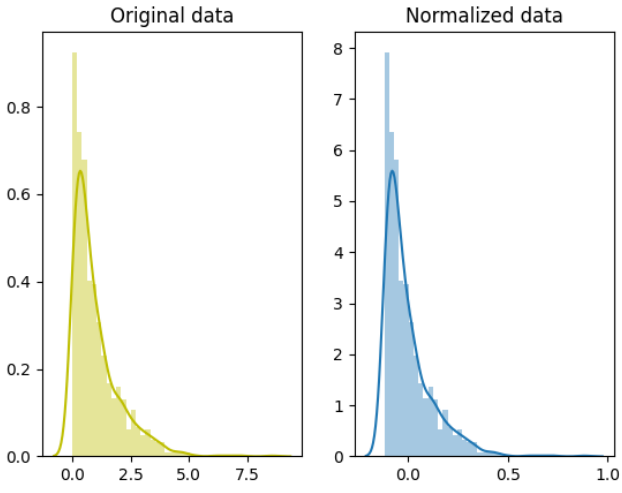

Fig. 1. Ejemplo de Normalización.

Para convertir cualquier rango a [0, 1], hay que restarle el valor más pequeño en el rango a cada uno de los valores. Y después, cada uno se debe dividir entre la resta del mayor y el menor valor. La fórmula se puede ver en Fig. 2.

$$\bar{x}^{(j)} = \frac{x^{(j)} - min^{(j)}}{max^{(j)} - min^{(j)}},$$

Fig. 2. Fórmula de Normalización.

Por ejemplo, si el rango de un feature es [350, 1450], hay que restarle 350, y dividirle 1100 a cada uno de los valores para obtener el rango de los valores [0, 1].

B. Estandarización

Se buscar transformar los valores para que los datos sigan una distribución normal. Recordemos que una distribución distribución normal es una distribución Gaussiana con $\mu = 0$ y $\sigma = 1$. El resultado de una estandarización se puede apreciar en Fig. 3.

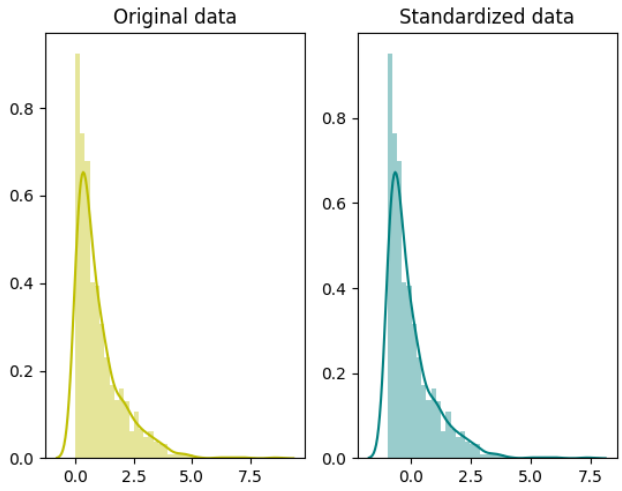

Fig. 3. Ejemplo de Estandarización.

Para ajustar un conjunto de datos a una distribución normal, a cada valor se le debe restar la media y después dividirlo entre la desviación estándar. La fórmula se puede ver en Fig. 4.

$$\hat{x}^{(j)} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}}.$$

Fig. 4. Fórmula de Estandarización.

Por ejemplo, si un feature de la estatura de varias personas tiene una media de 170 cm y una desviación estándar de 10 cm, se le debe restar 170 a cada dato y dividir el resultado entre 10.

C. Normalización vs. Estandarización

La estandarización es el método de escalado ideal cuando los datos ya siguen una distribución Gaussiana, cuando existen outliers para que estos valores no tomen mucho peso en el algoritmo, y también cuando se está utilizando un algoritmo de aprendizaje no supervisado.

Si los datos no siguen una distribución Gaussiana o no se quiere asumir la distribución de los valores, es mejor utilizar la normalización.

III. Redes Neuronales

Las redes neuronales son un método de inteligencia artificial en el que las computadoras aprender a procesar datos de una manera que está inspirada en la forma en que lo hace el cerebro humano. Utiliza nodos llamados neuronas que están en una estructura de capas para aprender de sus errores y mejorar continuamente. Una visualización de las capas de neuronas y sus conexiones se encuentra en Fig. 5.

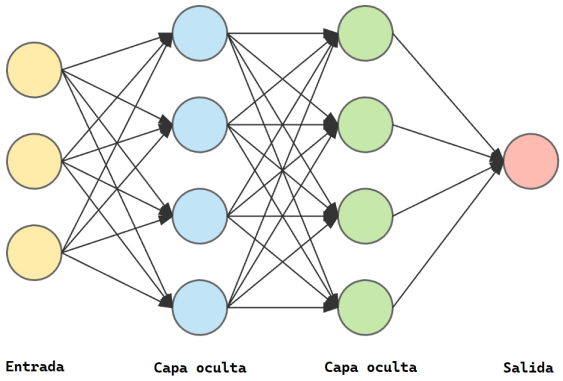

Fig. 5. Ejemplo de Redes Neuronales.

A. Dataset MNIST

MNIST es un dataset de imágenes de dígitos escritos a mano con el objetivo de reconocer el dígito que se encuentra en la imagen. Contiene un total de 70.000 imágenes, cada imagen es un dígito del 0 al 9, representado en una cuadrícula de 28x28 píxeles en escala de grises. Cada imagen es representada como una matriz de pixeles, que a su vez se puede aplanar a un vector y así cada pixel se comporta como un feature de la imagen (cada imagen tendría 784 features). Algunos ejemplos de las imágenes que incluye el dataset se puede ver en Fig. 6.

B. Regresión Logística

El problema de la clasificación de imágenes no se puede tratar como un problema de clasificación binaria ya que tenemos que clasificar las imágenes en 10 clases distintas (una para cada dígito). Pero podemos separar el problema en 10 clasificaciones binarias distintas: ¿Es un 0 o no?, ¿Es un 1 o no?, ¿Es un 2 o no?, ...

Para cada imagen se crea un vector con el resultado de cada clasificación binaria. Esto resultaría en un One-hot vector para cada imagen del dataset.

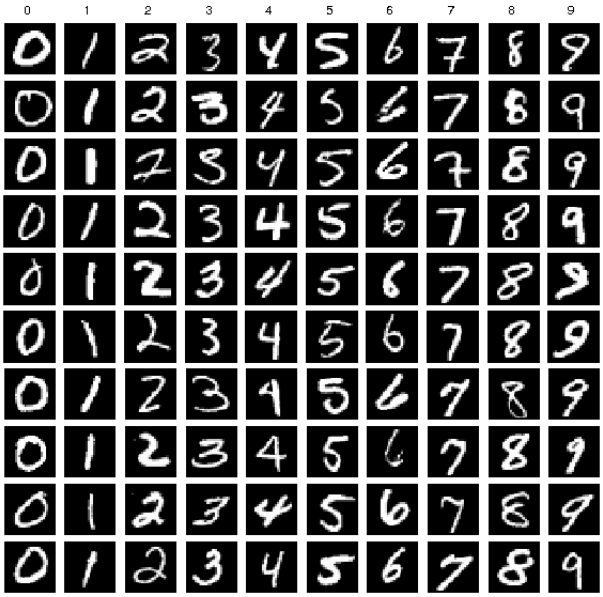

Fig. 6. Algunas de las imágenes en el dataset MNIST.

C. Matriz de Pesos

Para no hacer los cálculos de cada una de las regresiones logísticas con vectores, los resultados se pueden mostrar como una matriz. Esto simplifica los cálculos masivamente porque ya no hay que hacer n operaciones con cada vector, sino que solamente una con la matriz. Fig. 6.

En este caso, cada fila representa un solo pixel, mientras que las columnas son cada una de las regresiones logísticas.

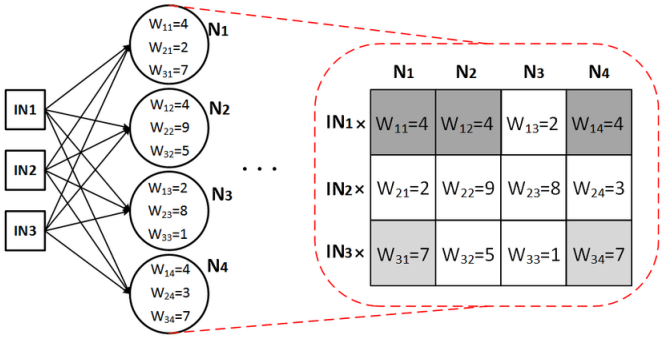

Fig. 7. Matriz de pesos.

REFERENCES

[1] J. Chong, "What is Feature Scaling & Why is it Important in Machine Learning?", 2020. [Online] Available at: https://towardsdatascience.com/what-is-feature-scaling-why-is-it-important-in-machine-learning-2854ae877048

[2] C. Liu, "Data Transformation: Standardization vs Normalization", 2023. [Online] Available at: https://www.kdnuggets.com/2020/04/data-transformation-standardization-normalization.html

[3] Amazon Web Services, "¿Qué es una red neuronal?". [Online] Available at: https://aws.amazon.com/es/what-is/neural-network/.