

Apuntes Semana 7 - 1/4/2025

Alvaro Vargas
Instituto Tecnológico de Costa Rica
Cartago, Costa Rica

I. MÉTRICAS

Son medidas que indican el rendimiento de un modelo. Se usan como formas objetivas de evaluar y compararlos. Para modelos de clasificación, como regresión logística, existen métricas específicas.

A. Matriz de confusión

Es una forma de visualizar los posibles resultados de un modelo de clasificación, donde las predicciones de las clases pueden ser correctas o no. Para la clasificación se pueden tener cuatro resultados:

- Verdadero positivo (T_p), de forma correcta se predice que sea cierto
- Verdadero negativo (T_n), de forma correcta se predice que no sea cierto
- Falso positivo (F_p), de forma incorrecta se predice que sea cierto, conocido como error estadístico tipo 1
- Falso negativo (F_n), de forma incorrecta se predice que no sea cierto, conocido como error estadístico tipo 2

Actual value	Positive	TP	FN
	Negative	FP	TN
		Positive	Negative
		Predicted value	

Fig. 1. Matriz de confusión

B. Accuracy

Mide la tasa de clasificaciones correctas sobre todas las clasificaciones hechas.

$$accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

C. Precisión

Mide la tasa de clasificaciones positivas correctas sobre todas las clasificaciones que se hayan dado como positivas, correctas o no. Es una forma de medir la cantidad de los

errores estadísticos de tipo 1 (falsos positivos); una precisión de 1 no tendría clasificaciones positivas incorrectas.

$$precision = \frac{T_p}{T_p + F_p}$$

D. Recall

Mide la tasa de clasificaciones positivas correctas sobre todos los valores que eran positivos, identificados correctamente o no. Es una forma de medir los errores estadísticos tipo 2 (falsos negativos). Entre más valores positivos que no clasificados correctamente, más bajo será el recall.

$$recall = \frac{T_p}{T_p + F_n}$$

E. F1-Score

Una métrica que para medir qué tan bien un modelo clasifica datos. Hace un balance entre las métricas de recall y precisión.

$$F1Score = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Del contexto dependerá la importancia de las diferentes métricas.

II. FEATURE SCALING

Cuando hay features con magnitudes de valores muy diferentes el modelo podría requerir ajustar la escala de estos valores. Para esto se pueden usar la normalización y la estandarización.

A. Normalización

Los features se ajustan a una nueva escala, normalmente de 0 a 1. Se puede usar la formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Donde x es el valor en la escala original, x' es el valor reescalado.

B. Estandarización

Hace que los nuevos valores tengan un promedio de cero y una desviación estándar igual a uno. Los valores puede calcularse con:

$$x' = \frac{x - \bar{x}}{\sigma}$$

Donde x' es el valor nuevo, \bar{x} es el promedio de x y σ es la desviación estándar.

C. Comparación

- Una no es mejor que la otra, escalan los features de formas diferentes.
- Si no se quiere asumir la distribución de datos es preferible normalizar.
- Cuando se ocupe una distribución Gaussiana o cuando hayan grandes outliers sera mejor estandarizar.

III. REDES NEURONALES

Las redes neuronales son modelos mas complejos formados por capaz de neuronas, cada neurona representa una función matemática.

Para dar un ejemplo se presenta el MNIST dataset. Con 60 000 imágenes de los números 0 al 9 dibujados, como se ven en la figura 2. Las imágenes pueden ser vectorizadas (flatten), así cada pixel representa un feature. Cada sample es de 28x28 pixeles, 784 features.

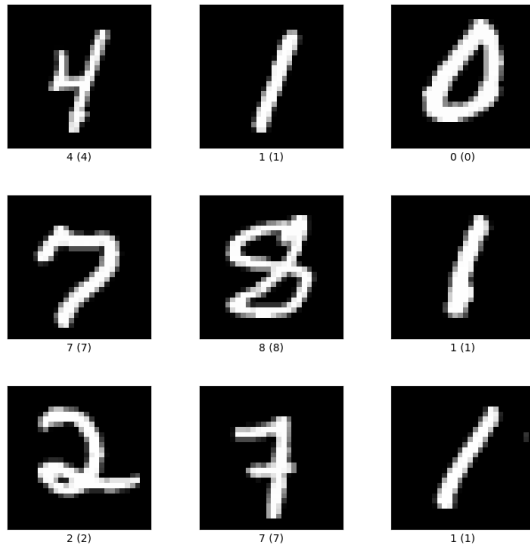


Fig. 2. Samples del MNIST dataset

Como podría ser el modelo que clasifica cada imagen en alguno de los diez números posibles (0-9)?

A. Múltiples regresiones logísticas

Podría existir una regresión logística que clasifique para cada uno de los diez posibles números. Cada una de estas regresiones sera una neurona. La misma imagen, usando su vector de features, se clasificaría en diez regresiones como se ve en la figura 3. Por cada regresión existiría un vector de pesos y un valor escalar de bias asociados. Los vectores de pesos se podrían combinar para formar una matriz de pesos, los valores escalares se podrían combinar para formar un vector de bias. Con esto se podría controlar las diez regresiones con una sola estructura y no una por una. A esta unión de neuronas se le llama capa.

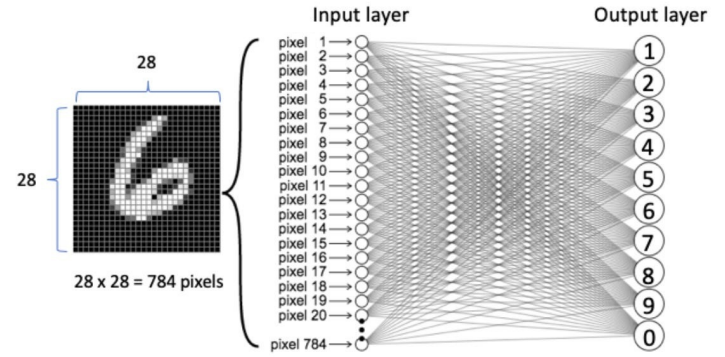


Fig. 3. Capa de entrada y capa de clasificación para MNIST

B. Capaz

Cada capa en una red neuronal debería ser diferenciable de otras capaz. Si la capa es derivable es optimizable. La salidad de una capa puede ser usada como la entrada de otra capa. Si en una capa hay no linealidad en la que le sigue se va a continuar esa no linealidad. Que este tipo de modelo puede trabajar con no linealidad hace que sea muy útil para resolver problemas complejos.