

Classification, Chapter 3

Robinson Daniel Mendoza Campo

Abril 2025

1 MNIST

El **dataset MNIST** es el considerado "Hello World" de la visión artificial. Contiene un conjunto de entrenamiento de 60.000 imágenes de dígitos manuscritos (de 0 a 9), y otro conjunto de pruebas con 10.000 muestras adicionales.

Las imágenes originales fueron normalizadas de forma que se pudieran contener en un grid de 20x20 píxels manteniendo las proporciones de la imagen original, y el resultado se centró en un grid de 28x28 píxels. Es de estas modificaciones de donde proviene la "M" de "MNIST" (Modified National Institute of Standards and Technology, Instituto de Estándares y Tecnología de los Estados Unidos).

Las muestras incluidas en el conjunto de entrenamiento fueron el resultado de escanear dígitos manuscritos de 250 personas (estudiantes de "high schools" y empleados de la oficina del Censo de los Estados Unidos). El dataset de pruebas contiene dígitos escaneados de otras 250 personas diferentes (con los mismos perfiles), lo que permite asegurar que los modelos obtenidos son capaces de interpretar dígitos incluso de personas no involucradas en la generación de los datos de entrenamiento.

2 Training a Binary Classifier

Un **clasificador binario** es un modelo de aprendizaje supervisado cuya función es asignar cada ejemplo de entrada a una de dos categorías mutuamente excluyentes (por ejemplo, "positivo/negativo", "spam/no-spam", "enfermo/sano").

- **Entrada:** un vector de características (numéricas o categóricas) que describen un objeto o instancia.
- **Salida:** una etiqueta binaria (0/1, -1/+1, verdadero/falso) que indica a cuál de las dos clases pertenece esa instancia.

En el entrenamiento, el algoritmo ajusta sus parámetros minimizando una función de pérdida (como la entropía cruzada o la hinge loss) sobre un conjunto de datos ya etiquetados, de modo que luego pueda generalizar y predecir correctamente la etiqueta de nuevos datos.

3 Performance Measures

- **Exactitud (Accuracy)**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- **Precisión (Precision)**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Exhaustividad / Sensibilidad (Recall)**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Puntuación F1 (F1-score)**

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Área bajo la curva ROC (AUC-ROC)**

$$\text{AUC} = \int_0^1 \text{TPR}(t) d(\text{FPR}(t))$$

donde $\text{TPR}(t)$ y $\text{FPR}(t)$ son la tasa de verdaderos positivos y la tasa de falsos positivos al umbral t .

4 Multi-class Classification

La **clasificación multiclase** es una tarea de aprendizaje supervisado en la que un modelo asigna cada observación de entrada a una única etiqueta dentro de un conjunto de K categorías mutuamente excluyentes (con $K > 2$). A diferencia de la clasificación binaria (dos clases), aquí el objetivo es aprender una función

$$f : R^d \longrightarrow \{1, 2, \dots, K\}$$

que, dada una nueva muestra $\mathbf{x} \in R^d$, prediga correctamente a cuál de las K clases pertenece.

- **Entrada:** vector de características $\mathbf{x} = (x_1, x_2, \dots, x_d)$.
- **Salida:** etiqueta $y \in \{1, 2, \dots, K\}$.

5 Error Analysis

6 Multi-label Classification

La **clasificación multilabel** (o *multilabel classification*) es una tarea de aprendizaje supervisado en la que a cada instancia de entrada se le pueden asignar simultáneamente varias etiquetas de un conjunto de K posibles. A diferencia de la clasificación multiclase, donde cada ejemplo pertenece a una única clase, aquí una misma muestra puede tener cero, una o varias etiquetas activas al mismo tiempo.

Definición formal

Sea $\mathbf{x} \in R^d$ un vector de características y $\mathbf{y} \in \{0, 1\}^K$ un vector binario de etiquetas, donde

$$y_k = \begin{cases} 1, & \text{si la instancia pertenece a la etiqueta } k, \\ 0, & \text{en caso contrario.} \end{cases}$$

El objetivo es aprender una función

$$f : R^d \longrightarrow \{0, 1\}^K$$

que, dada una nueva muestra \mathbf{x} , prediga para cada etiqueta k si está presente (1) o no (0).

7 Multi-output Classification

La **clasificación multi-output** (o *multi-target classification*) es una extensión de la clasificación supervisada en la que el modelo debe predecir simultáneamente varias variables de salida, cada una de las cuales es una tarea de clasificación por separado. A diferencia de:

- **Clasificación multiclase**, donde hay una única etiqueta

$$y \in \{1, \dots, K\}.$$

- **Clasificación multilabel**, donde se predice un vector binario

$$\mathbf{y} \in \{0, 1\}^K$$

indicando la pertenencia a varias clases de un mismo problema.

En *multi-output* trabajamos con múltiples variables de salida distintas. Por ejemplo:

- Para cada imagen, predecir tanto el objeto principal (10 clases posibles) como el color predominante (5 clases posibles).
- En un diagnóstico médico, asignar simultáneamente la presencia/ausencia de varias enfermedades (varias salidas binarias) y, además, clasificar su gravedad (varias salidas multiclase).