

Índice

1. MNIST (Modified National Institute of Standards and Technology database)	1
1.1. Enlaces	1
1.2. Descripción general	1
1.3. Estadísticas resumidas	1
1.4. Subdivisión de datos	1

1. MNIST (Modified National Institute of Standards and Technology database)

1.1. Enlaces

- Wikipedia: MNIST en Wikipedia

1.2. Descripción general

MNIST es un conjunto de datos de imágenes de dígitos escritos a mano, ampliamente utilizado como benchmark en aprendizaje automático y visión por computadora. Contiene imágenes grises etiquetadas del 0 al

1.

1.3. Estadísticas resumidas

- Número total de ejemplos: 70.000
- Número de clases: 10 (dígitos 0–9)
- Número de píxeles por imagen: $28 \times 28 = 784$
- Rango de valores de píxeles: 0–255

1.4. Subdivisión de datos

- Entreno: 60.000 imágenes
- Validación: 10.000 imágenes

Aunque habitualmente se toman las últimas 10.000 imágenes como conjunto de validación, el muestreo original no fue completamente aleatorio: los

dígitos provienen de diferentes fuentes y autores y se mantiene cierto orden de recopilación, por lo que la distribución no es homogénea. Por ello, es recomendable **barajar** los datos antes de subdividirlos en conjuntos de entreno y de validación, para evitar sesgos.