

Clase ejercicios - IMA539.

Eduardo Castro Thompson.
Estudiante Ingeniería Civil Física.
June 30, 2023.
Pauta.

La base de datos.

El conjunto de datos Custom-FashionMNIST contiene un total de 70.000 imágenes en escala de grises de tamaño 28x28 píxeles cada una. Las imágenes provienen de artículos publicados en Zalando se distinguen 10 clases:

Clase	Etiqueta	Clase	Etiqueta
0	Polera	5	Sandalia
1	Pantalón	6	Camisa
2	Polerón	7	Zapatilla
3	Vestido	8	Bolso
4	Abrigo	9	Botas

Este dataset es importante en la literatura pues se diseñó para tomar el puesto del MNIST original (clasificación de números escritos a mano). Normalmente, su carga se puede automatizar en la mayoría de las librerías de ML como PyTorch o Tensorflow, sin embargo, Custom-FashionMNIST ha sido reordenado cuidadosamente de modo de hacerle el trabajo un poco más difícil:

El conjunto de datos se ha separado de modo que los elementos clasificados con un número

- **Par:** Se encuentran separados en carpetas cuyo nombre es el número de la clase y cada archivo de imagen posee como nombre `image-id_A.jpg`, donde `A` es el identificador de la imagen. La única excepción a la regla son los elementos con clasificación bolso, en cuyo caso el nombre de la carpeta es un string.
- **Impar:** Se encuentran todas agrupadas en un directorio de nombre `other` y el formato de los nombres de los archivos de imagen es `image-id_A-label-B.jpg`, donde `A` es el identificador de la imagen y `B` su clase.

Los identificadores `A` de las imágenes son únicos, independientemente de su ubicación o clasificación.

El siguiente diagrama muestra la distribución de los ficheros del conjunto de datos:

```
Custom-FashionMNIST
├── 0
│   ├── image-id_0.jpg
│   └── ...
├── 2
│   ├── image-id_5665.jpg
│   └── ...
├── 4
│   ├── image-id_1837.jpg
│   └── ...
├── 6
│   ├── image-id_31784.jpg
│   └── ...
├── bag
│   ├── image-id_18461.jpg
│   └── ...
└── other
    ├── image-id_6171-label_1.jpg
    └── ...
```

Importante: Como el sistema operativo en que este conjunto de datos fue armado es MacOS, carpetas con el nombre `.DS_Store` pueden ser encontradas dentro de algunos directorios, lo cual debe ser considerado en la lectura de los archivos.

Ejercicio.

Realice un split train/val/test de la base de datos, luego, idee e instancie un modelo de CNN utilizando PyTorch. Entrene el modelo y determine si este presenta sobreajuste, bajo-ajuste. Finalmente, reporte Acc, TPR, FPR, K y MCC en el conjunto de testeo.