

03 - Handle data and preprocessing

Maxime Ellerbach

March 2023

Can be pretty much anything

- Scalar
- Image
- Vector
- Sequences
- Categories
- ...

Preprocessing and encoding

In order for the model to understand the data, it must be encoded in a certain way

- floats (could also be int8 depending of the quantization)
- usually normalized between 0.0 and 1.0

Preprocessing and encoding - Classification example

Let's consider the following input data:

- Image of a number - shape: (28,28,1)

output data (Label):

- class of the number - 0, 1, 2, ... 9

How to encode this into something the model could understand ?

Preprocessing and encoding - Classification example

- Image : normalize the pixels (if not already) then transform the 3D array into 1D - from $(28,28,1)$ to (784) . We could imagine to have more data preprocessing: apply some filters, ...
- Label : transform the number into a one-hot encoded array

One hot encoding

- Dataset
 - Image_of_2 - [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
 - Image_of_9 - [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
 - ...
 - Image_of_6 - [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
- Output shape of the model is then: (10,)
- During the inference process, we take the maximum value predicted eg: [0, 0, 0, 0.72, 0, 0, 0, 0, 0.28, 0], would mean the model predicted most likely a 3.

Practice 01 - Data Preprocessing