

# **Machine Learning: SVM for Classification, NN for Classification, and Clustering**

Fabio Vandin

December 7<sup>th</sup>, 2021

# Instructions

You have to complete 2 Jupyter notebooks:

- one for SVM on classification (digits images data - MNIST)
- one for NNs on classification and clustering (clothes images data)

Both notebooks have missing code: need to fill in what is missing

You also need to write some text (to explain choices or describe results)

Feel free to add cells with text if you need to explain or describe some ``non-standard'' decision!

**FIRST THING TO DO: you need to put your ID number in both notebooks (as seed for random number generators).**

**IMPORTANT: code already there is a guideline, if you want to change or rewrite some/all of it go ahead, but make sure to answer all TO DOs!**

# Deadline

Submit your completed notebooks:

- **deadline: Thursday, January 20<sup>th</sup>, 11:55 PM**
- **use link in elearning website**

Submit 2 files (1 completed notebook for SVM on classification, 1 completed notebook for NNs on classification and clustering) - **Only submit your completed notebooks!**

**IMPORTANT:** Use the following file names for the 2 files that you have to submit:

- for the SVM classification notebook: SVM\_FirstnameLastName\_IDnumber.ipynb
- for the NNs classification notebook and clustering:  
NN\_FirstnameLastName\_IDnumber.ipynb

**Example:** student Fabio Vandin, ID number 000001 will submit files:

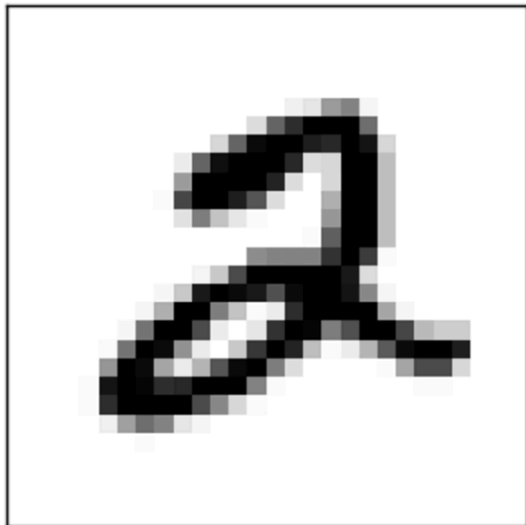
- SVM\_FabioVandin\_000001.ipynb
- NN\_FabioVandin\_000001.ipynb

# Dataset for SVM classification

MNIST dataset:

- <http://yann.lecun.com/exdb/mnist/>
- <https://www.openml.org/d/554>
- 70000 training instances, input =vector of 784 integer values in  $[0, \dots, 255]$  = 28x28 matrix

INPUT:

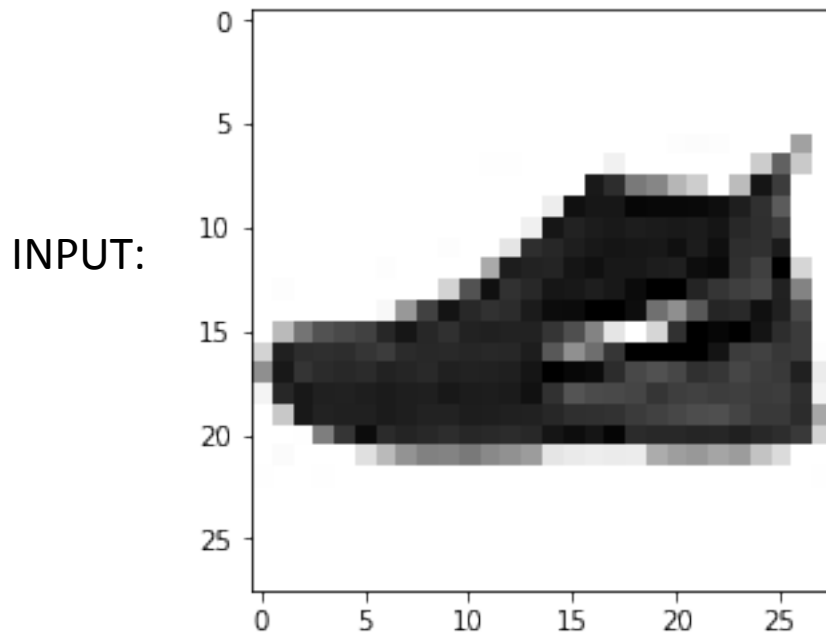


TARGET/LABEL: 2

# Dataset for NN classification and Clustering

Fashion MNIST dataset:

- see this link for more info:  
<https://pravarmahajan.github.io/fashion/>
- 60,000 training instances, input = vector of 784 integer values in  $[0, \dots, 255]$  = 28x28 matrix
- Instance = image from one of 10 clothes categories



TARGET/LABEL: 7  
(= sneaker)

# Get the data for NN classification and Clustering

Download the following file (26MB):

<http://www.dei.unipd.it/~vandinf/courses/ML2021/data.zip>

Unzip the file so that the “data” folder is within the “ML2021\_HW2\_and\_HW3” folder you downloaded from elearning.