PHYS 243

Homework 4

Deadline: May 24, 2021, at 11:59 AM

Classify the Handwritten digits!

In this homework set you are going to use the MNIST handwritten data set. You can load it using the following:

```
# Loading the dataset
from sklearn.datasets import fetch_openml
MNIST = fetch_openml('mnist_784', version=1, cache=True)
X = MNIST.data # Data with pixels values
y = MNIST.target # Array of labels
```

The task here is to classify the digits into their own category.

Find all the 9s!

In this section you should build up a classifier that can distinguish number 9 from every other numbers. (reusing code and libraries are ok as long as you explain what is going on)

For each section below you need to measure your performance. So, make sure to run the performance check at every part.

1. Find the 9s using Logistic regression.
2. Find the 9s using K-Nearest neighbours for Minkowski metric of order (1, 2, 3).
3. Find the 9s using Support Vector Machines.

Find the clusters in MNIST

Using the k-means algorithm find the k clusters (here is obviously 10 clusters) in your data. Find what fraction of the labels in each cluster are the same as the majority label within that cluster.

Find every single digits!

1. Find the digits using Logistic regression and one-vs-rest for multi-class option.
2. Find the digits using K-Nearest neighbours for Minkowski metric of order (1, 2, 3).
3. Find the digits using Support Vector Machines and one-vs-rest for multi-class option.
4. Comment on any significant difference between your results for the **binary classifier** vs **multi-class classifier**s.

Homework 4

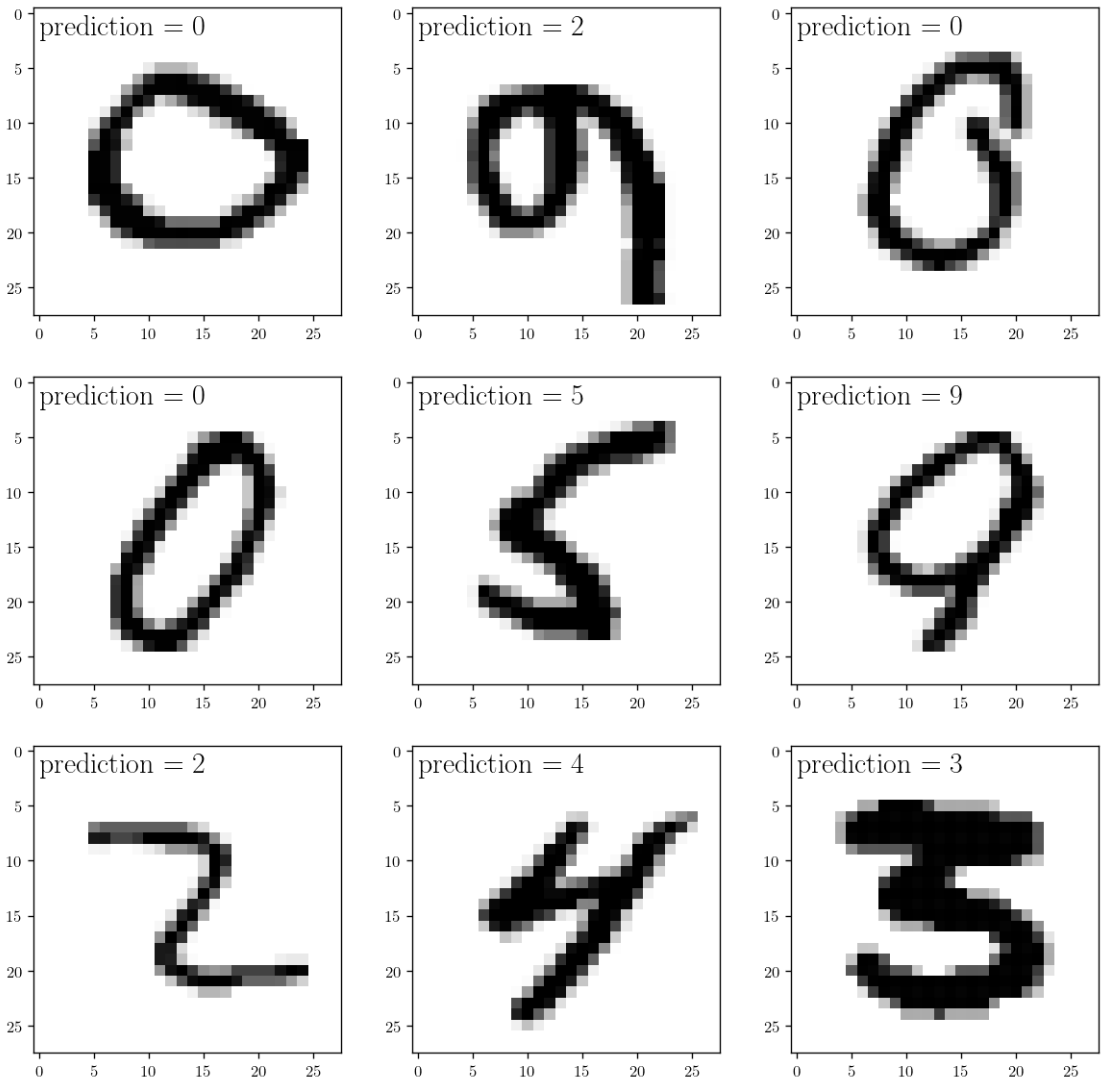

Figure 1: Sample from the MNIST dataset and SVM predictions