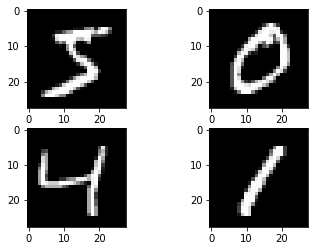
MNIST Project Report

Yusif Imamverdiyev M1DSAI

Introduction to data science

Project name : **MNIST [hard]**

The dataset is a collection of 70.000 images of handwritten digits. First i load the data using scikit-learn provides fetch\_openml convenience function to download the data set if is not found on disk and read it into an object. Some samples given bellow.

  
The MNIST data set is partitioned into a trainig set of 60.000 images and test set of 10000 images. The dataset is commonly used to evaluate a variety of machine learning models. It is popular because little preprocessing is required . Let’s use scikit learn to build a classifier that can predict the digit depicted in an image

First i imported all necessary classes

from sklearn.pipeline import Pipeline

from sklearn.preprocessing import scale

from sklearn.model\_selection import train\_test\_split

from sklearn.svm import SVC

from sklearn.model\_selection import GridSearchCV

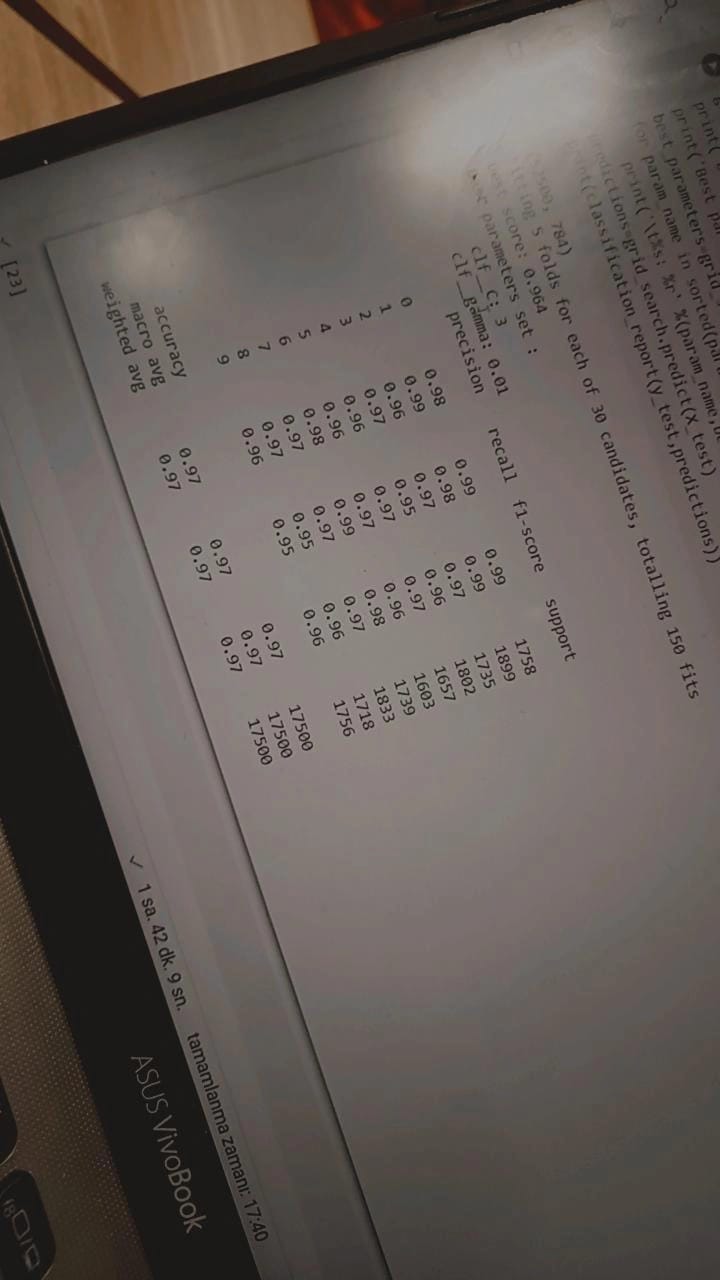
from sklearn.metrics import classification\_report

Then i load my data set with convenience function again . I scale features and center each feature around the origin. We then split the preprocessed data into training and test using

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,random\_state=11)

Next i instantiate an SVC or support vector classifier, object. This object exposes an API like that of scikit-learn’s other estiomators. The classifier is trained using the fit method and predictions are made using predict Method.

The mods interesting hyperparameters of SVC are set bu the kernel,gamma and c keyword arguments. The kernel keyword arguments specifies the kernel to be used. Scikit-learn provides implementations of the linear,polynomial,sigmoid,and radial basis function kernels . The degree argument should also be set when teh polynomial kernel is used. C controls regularization .Gamma is kernel coefficient for the sigmoid . And our Final output is the following.



I do not want to use this photo which i take by my phone . But for running this program took more time . And when i writing this report i have no a lot of time 😊

The best model has an average F1 score of 0.97 . This score can be increased Further by training on more than the first then thousand instances