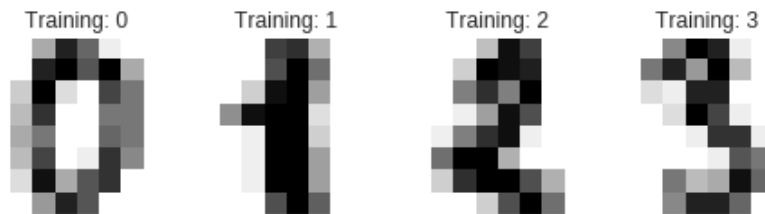


Handwritten Digits Classification

Handwritten Digits Classification is an interesting topic of Machine Learning. And it is one of the most popular demonstrations of the capability of deep learning techniques. In this report, the sample data is composed of plenty of 8 x 8 pixel images of digits. I tried to train classifiers to recognize these digits from images, while used data exploration and data visualization to explain why each model works or does not. I used matplotlib, sklearn, and yellowbrick packages for analyzing the data.

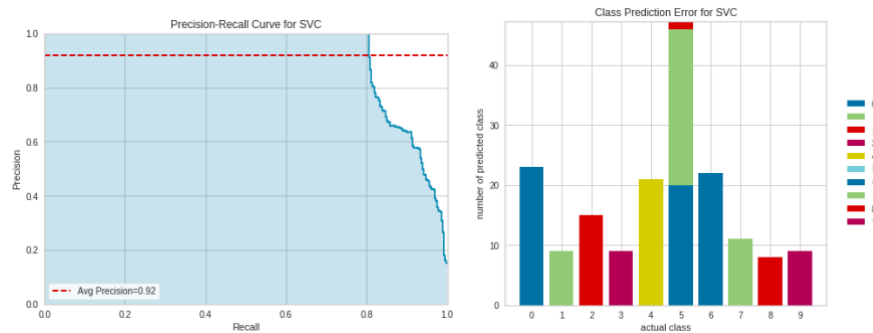
Firstly, I utilized datasets to import the data into the Colab. And then, I took a look at the first 4 images (shown as below) which stored in the 'images' attribute of our dataset.



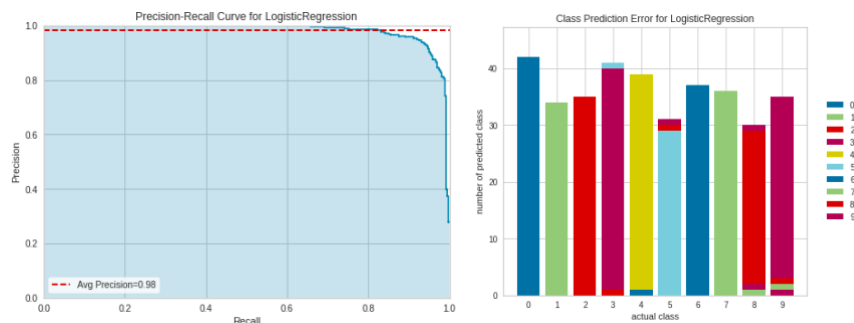
In order to modeling and training, I split the data into two parts: the train data set and the test data set. And the test data set is 20% of the raw data.

Later, I utilized precision-recall curves and Class Prediction Error to figure out which model is the most suitable for the digits recognition. The following are the results of visualization of precision about five models, including the SVC, Logistic Regression, Ridge Classifier, Random Forest Classifier, and K-nearest Neighbors Classifier.

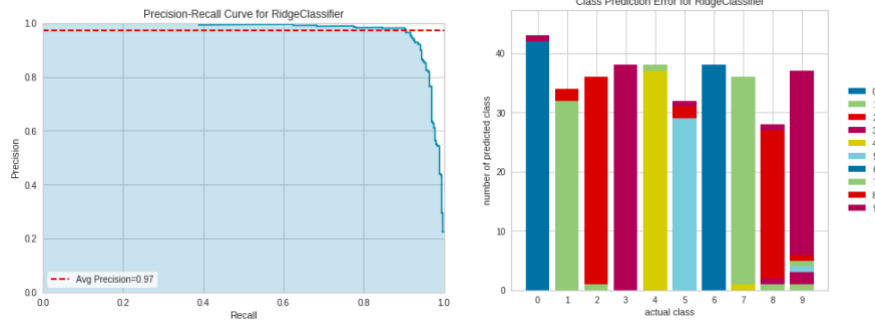
SVC:



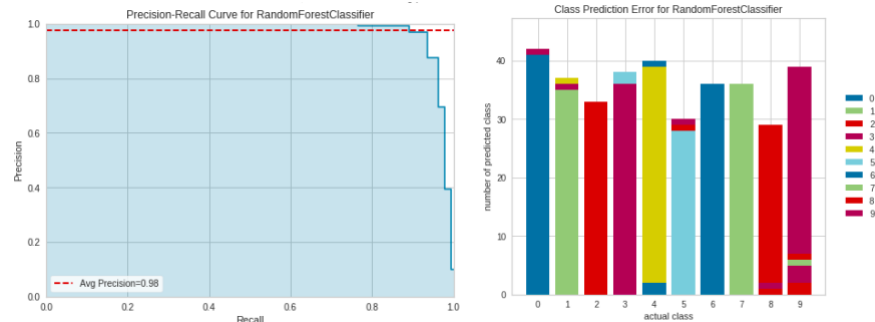
Logistic Regression:



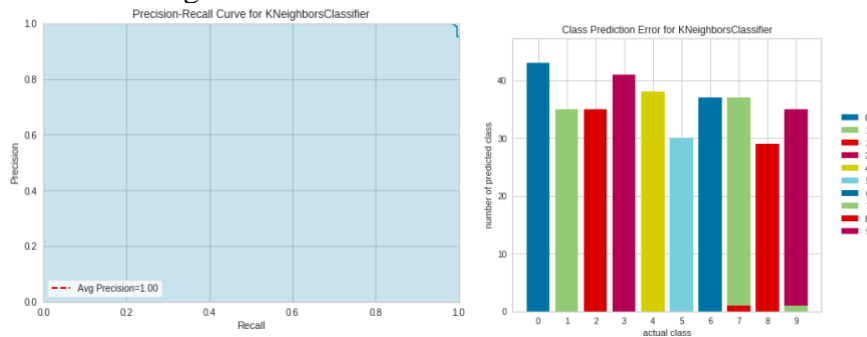
Ridge Classifier:



Random Forest Classifier:



K-nearest Neighbors Classifier:



(Knn: the precision-recall curve is almost made up by the blue, which means the precision is almost equal to 1, and the class prediction showed there is only a few of error within each class. Hence, this is the best model among these five models)

In conclusion, according to above visualization (left: Precision-recall Curve, right: Class Prediction Error), we can easily figure out the K-nearest Neighbors Classifier is the most precise model to deal with digits recognition, and the SVC is the worst one. Hence, when we want to deal with the digits classification problem, we can try k-nearest neighbors firstly.

My Colab Link:

<https://colab.research.google.com/drive/117ojZ-HjG6cynDL-GITQTTG5ElbX76>