Problem Statement:

Currently label the MNIST dataset to identify the corresponding numbers the image corresponds

Solution Approach

Data Exploration

1. Take 10000 rows for training.
2. Add a new feature – ratio of black and white pixels for each available label
    a. The classification of the pixels is done using exploratory search to identify maximum data variance for one threshold
    b. Identify average value of the ratio of black and white for each label
    c. Append values to main dataset

Data Preprocessing

1. Scaling of predictor variables using sckitlearn.scaler()
2. Test – train at 7:3 ratio

Data Modelling

1. Linear Model – accuracy of 93%
2. RBF kernel based SVM model –
    a. Hyperparameter tuning on 5 folds
    b. Range of C, gamma given
3. Final SVM RBF model is modelled with C=15, gamma=0.001 : accuracy of 95%