Report on MNIST Classification

Feature Extraction and reasoning:

In the given case study we have been given 28*28 cross size images in the form of a 1-D list. As we know each pixel in any image works as a feature. As far as classical machine learning is concern, we can directly feed these features to the model. In the case of this problem, we try to extract the feature using one of the widely used CNN model for feature extraction i.e. VGG16. We used CNN because CNN based features work really well for image data. But when we train and validate our model on the extracted features using SVM and XGboost, the performance was very bad so we are not using extracted features for the modeling.

What classification techniques did you try?

As our model was having a large number of features, so to avoid the curse of dimensionality, I used Logistic Regression and XGboost for the image classification. The XGboost in general works really well. In our problem, XGboost performed well as compared to Logistic Regression.

Which of the methods (and for what hyperparameters) showed the best cross-

XGboost model performed well as compared to Logistic Regression
Optimal Set of Hyperparameters:
Optimal Depth = 100
N_estimators = 150
Validation Accuracy =.81

What test accuracy are you expecting?

I am expecting test accuracy to be more than 80%.