Report on Stack Ensembles

Sumit Chawan

In each stacking technique, logreg was used as the stack classifier while the base classifier was trained using svm, tree, logreg.

The below dataset were used to test the performance of the classifier: -

- Test on Validation Set
- Test on test set
- Test on entire MNIST test data file.

The following models were designed: -

- StackEnsembleClassifier :- The entire training set was used in order to train the stack layer.
- StackEnsembleHoldOutClassifier : In this we train the dataset on part of the training set and the remaining validation set is used in the stack layer.
- StackEnsembleClassifierKfold:- This classifier has the maximum complexity as the classifier is trained on k folds.
- StackEnsembleOneVsOneClassifier:- This classifier is trained on the combination of Unique output labels.
- Decision Tree and Bagging were used along with pruning to find best parameters which were compared to other approaches.

Model Performance

- After comparing all the above models, it was observed that Stacking ensemble with Kfold has the highest accuracy as compared to all the other approaches. Graphs are shown to indicate the accuracy.
- Decision tree and bagging even with best parameter approach yielded less performance as compared to Stacking ensemble with k fold.

Computational Complexity

The computational complexity of Stack Ensemble with One Vs One is more efficient as compared to other stacking models while Stacking ensemble with k fold took the maximum time to fit.

Model Complexity

Stacking ensemble with K fold has the highest complexity as compared to its counterparts as it builds the model by iterating over k folds.