Assignment on Classifiers

Due Date: 25th April 2020

Question

Use the Sklearn’s wine dataset and perform the following tasks. Use train test split (Scikit-learn’s) to make a stratified split of 70-30 train-test with seed 42.

* 1. [3 pts] Plot pairwise relations in dataset using seaborn and give inferences.
  2. [5 pts] Implement Gaussian Naive Bayes using scikit-learn and report all evaluations and any other observations that you may have.
  3. [7 pts] Use Scikit-learn’s DecisionTree and tune the hyper-parameters(Height/splitting criteria) balancing the run time and accuracy. Report all evaluations and observations.
  4. [5 pts]  Implement KNN using scikit-learn and report all evaluations and any other observations that you may have.
  5. [5 pts] Compare the above four models (Gaussian Naive Bayes, KNN and Decision Tree) on all evaluations and training time. Defend the best model for the dataset used.

Evaluations should be in the form of F-1 score, Accuracy and ROC Curve (You can use sklearn for all evaluation metrics). Additional observations may include analysis showing evidence that the model has trained correctly, class-wise accuracy.