CENG 463 Homework 2 due to 31th October 2022 at 13.00.

Analyze Confusion Mat and Compare Performance Metrics for Different Classifiers.

Description:

We are going to explore classification metrics with different classifiers and their implications to Spam e-mail classification.

You may use other sklearn function such as accuracy and train/test data splitting.

Start with the Lecture 4 Jupyter Notebook

Tasks:

1. Classify KNN with K=1, SVM (from Week 3's lab), and Naïve Bayes
2. Analyze Confusion Matrices for each classifier for how many samples are misclassified as FP, FN. What are the consequences. Discuss your analyzes.
3. Compare Precision, Recall and F-Score/F-measure against Accuracy which one is useful to detect missed e-mail or which one you would prefer to classify for this task? Would you use this e-mail service? Another discussion is expected. What does precision and recall metrics imply?
4. Plot ROC curves and briefly discuss its shape for comparison of classifiers.

Notes:

Measure and report computation times and classification accuracies in your notebooks.

You may group with another student, teams of 2 students are allowed.

Only your notebooks for your group will be submitted. Add Student ID and your initials such as MT1223456789_SecondStudentInitialsSecondStudentID.ipynb, put full names & IDs in the top of the file. The file and data are available in course GDrive share.

Add code comments and discuss your findings in the notebook.

Elegant code, useful comments and efforts will be graded handsomely.

Good Luck

Dr. Mustafa Teke