

CMP 466 – Machine Learning and Data Mining- Spring 2021

Team Project- Assignment 5

Please submit a report including the following by the deadline:

- 1- [10 points] Apply feature scaling using MinMaxScaler to scale your features.
- 2- [20 points] Pick the best two classifiers that you tried on your dataset so far. Using either cross-validation or random splitting, perform classification using each of the classifiers with the **scaled features as your input** to the classifiers. Report the following performance metrics: training time and accuracy, testing time and accuracy, precision, recall, F-score.
- 3- [10 points] Apply feature selection, using either SelectKBest or SelectPercentile, to select the best features to use.
- 4- [20 points] Considering the best two classifiers that you tried on your dataset so far. Using either cross-validation or random splitting, perform classification using each of the classifiers with the **selected features as your input** to the classifiers. Report the following performance metrics: training time and accuracy, testing time and accuracy, precision, recall, F-score.
- 5- [10 points] Apply PCA on your dataset, print the explained variance by each of the principle components (PCs) and select the first few components that explain most of the variance 70%+ (cumulative).
- 6- [20 points] Considering the best two classifiers that you tried on your dataset so far and using either cross-validation or random splitting, perform the classification using each of the classifiers with the **first few PCs that explain most of the variance as your input** to the classifiers. Report the following performance metrics: training time and accuracy, testing time and accuracy, precision, recall, F-score.
- 7- [10 points] Summarize and discuss your results for each of the classifiers. For example, discuss the impact of feature scaling, selection and dimensionality reduction on your results and running time. Estimate the improvement quantitatively.

Submission: Submit your assignment to iLearn by the deadline. Submission should include the report and the code. Late submissions will be penalized according to the syllabus late policy. No submissions will be accepted after four days of the deadline.