Introduction to Machine Learning
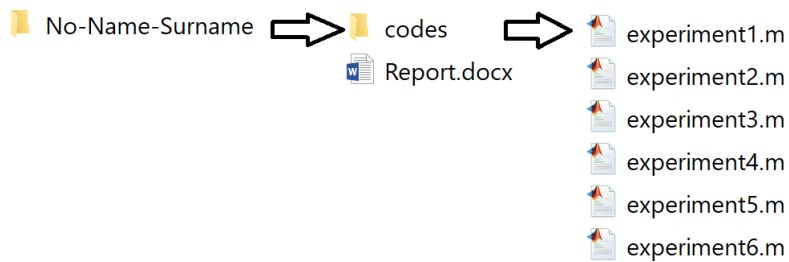
Midterm Homework

Due Thursday, November 26, 2020

You should create your report format.

Send report and all experiment MATLAB codes to machinelearning.kou@gmail.com

Your files should be as below:

No-Name-Surname ⟹ codes ⟹ experiment1.m
Report.docx   experiment2.m
     experiment3.m
     experiment4.m
     experiment5.m
     experiment6.m

- Use fishiris and ionosphere datasets for all experiments (sample datasets in MATLAB)

1) **Experiment 1**

- ➢ Visualize data

- ➢ Split the datasets randomly into a training (80 %) and a test set (20 %). Using linear SVM, report accuracy, precision, and recall, F1-score, TPR, FPR.

- ➢ Scale datasets using standardization: Compute the mean and standard deviation of each feature (i.e., over a column in the training set). Then subtract the mean from each value and divide each value by the standard deviation. 1). Then, split the datasets into a training (80 %) and a test set (20 %). Using linear SVM, report accuracy, precision, and recall, F1-score, TPR, FPR.

2) **Experiment 2**

- ➢ Split the datasets randomly into a training (80 %) and a test set (20 %).
  Use linear SVM, Decision Tree, kNN methods
  Report accuracy, precision, and recall, F1-score, TPR, FPR.
  For each classification method calculate confusion matrix

3) **Experiment 3**

- ➢ Use linear SVM, Decision Tree, kNN methods
  Implement k-fold Cross-Validation. Calculate precision and recall values for k=5,6,7,8,9,10 and plot these values.

4) Experiment 4

➢ Use linear SVM, Decision Tree, kNN methods
Split the datasets randomly into a training (80 %) and a test set (20 %).
Plot the receiver operating characteristic (ROC) curve for the classifiers (score values can be use as parameter). 2 ROC curves -> precison-recall and TPR-FPR

5) Experiment 5

➢ Use kMeans method
Split the datasets randomly into a training (80 %) and a test set (20 %)
Report accuracy, precision, and recall, F1-score, TPR, FPR
Visualize data and cluster centers
➢ Use kMeans method
Split the datasets randomly into a training (80 %) and a test set (20 %)
Calculate precision and recall values for k=1,2,3,4,5,6,
For all k values visualize data and cluster centers

6) Experiment 6

➢ Split the datasets randomly into a training (80 %) and a test set (20 %)
For each dataset calculate training time and prediction time for SVM, Decision Tree, kNN and kMeans
Give these values with a table as below

	Method-1	Method-2	Method-3	Method-4
F1-score				
Training Time				
Prediction Time				

Plot F1-score and prediction time for all methods in same figure as below