COEN 129 Machine Learning and Data Mining

Homework 2 (100 points)

Due: 3:30pm, Monday, April 24, 2017

Implement Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). You can use Python, R or Matlab for this assignment. Please do not use any machine learning library for this assignment. Perform classifications on the Iris dataset which can be downloaded at http://www.cse.scu.edu/~yfang/coen129/iris.data

The dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant.

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
   -- Iris Setosa
   -- Iris Versicolour
   -- Iris Virginica

Exercises:

1. Break the sample into 80% for training, and 20% for test datasets. You can choose the first 80% instances from each class for training and the rest for testing.
2. Build an LDA classifier based on the training data. Report the training and test errors for your classifier.
3. Build a QDA classifier based on the training data. Report the training and test errors for your classifier.
4. Is there any class linearly separable from other classes? Explain your answer based on your experiments.
5. Are any of the variables not important in classifying iris type? Explain your answer based on your experiments.
6. Assume the features are independent, i.e., $\sum$ is a diagonal matrix. Repeat 2 and 3, and report your results.