ECE/MSIM 607 Machine Learning I

Spring 2017, 4:20 pm - 7:00 pm R

INSTRUCTOR: Jiang Li, Associate Professor, ECE Department.

Office: 231D KH, Phone: 683-6748, Email: JLi@odu.edu

Office Hours: TR 10:00-11:30AM (Other times by appointment)

Project 1

**Due on Feb. 21st. 20 points penalty for the late project per day. Do not use Matlab build-in classifier functions. You need to implement those classifiers by your own. Other built-in functions like cov() inv() are allowed.**

You will design a Bayes classifier, a Naive Bayes classifier and a *k*-nearest neighbor classifier in this project. Attached are two data sets. One was generated (training and testing) and another one is for zip code recognition, which also contains training and testing sets. For the generated data, it is an 400 by 3 matrix for training and testing, respectively. Each row in the data sets is one observation. All observations belong to one of the two classes (class 0 and class 1). The first two columns are 2 dimensional observations and the third column contains class IDs. For the zip code datasets, there are 16 inputs (features) and the last column are the class IDs, which are from ‘0’ to ‘9’.

**Task 1:** For the generated datasets, scatter plot the training and testing data sets. Show class 0 in red and class 1 in blue.

**Task 2:** For both the generated data and the zip code data, design a Bayes classifier assuming that the data follows a Gaussian distribution. Estimate corresponding parameters from the training data (parametric estimation). Apply your Bayes classifier to the training and testing data sets, respectively. Report training and testing classification accuracies. (For the zip code dataset, if the inverse of any covariance matrix does not exist (this can be done by checking if the determinant of the covariance matrix is zero), then add a small positive value to all the diagonal components and report the value you added in your project report).

**Task 3:** Repeat Task 2 by designing a naive Bayes classifier. (A simple way to design a naïve Bayes classifier is to make those off-diagonal elements in the estimated covariance matrix zero.)

**Task 4**: Repeat Task 2. Utilize a nonparametric estimation technique to estimate the conditional distribution *p*(**x**|*C*i), using a Gaussian kernel. Try 3 different kernel sizes *h* at your choices. Report classification accuracies on the testing data set.

**Task 5**: For both the generated data and the zip code data, design a *k*-nearest neighbor classifier. Try 3 different *k* values at your choices. Keep in mind that *k* must be odd. Report classification accuracies on the testing data set. Turn in all of your codes and results.

**Task 6**: For the generated data set, display the decision boundaries produced by each of the above classifiers, using the method we have discussed in class. Show them separately.

Turn in all of your codes and results.