Project

CSE 802 - Pattern Recognition and Analysis
Instructor: Dr. Arun Ross
Due Date: May 2, 5:00pm

Note: You are permitted to discuss this project with the others in class, but the final report should reflect your own effort. Any indication to the contrary will be considered an act of academic dishonesty. A hard-copy has to be turned in. Also, a zipped version of your submission containing the code and the report must be sent via email to rossarun at cse.msu.edu by 5:00pm on 5/2/2014.

---

1. [65 points] Consider the dataset available here. It consists of two-dimensional patterns, $\boldsymbol{x} = [x_1, x_2]^t$, pertaining to 3 classes $(\omega_1, \omega_2, \omega_3)$. The feature values are indicated in the first two columns while the class labels are specified in the last column. The priors of all 3 classes are the same. Randomly partition this dataset into a training set (70% of each class) and a test set (30% of each class).

   (a) Let

   $$\begin{aligned} p([x_1, x_2]^t|\omega_1) &\sim N([0,0]^t, 4I), \\ p([x_1, x_2]^t|\omega_2) &\sim N([10,0]^t, 4I), \\ p([x_1, x_2]^t|\omega_3) &\sim N([5,5]^t, 5I), \end{aligned}$$

   where I is the 2×2 identity matrix. What is the error rate on the test set when the Bayesian decision rule is employed for classification?

   (b) Suppose $p([x_1, x_2]^t|\omega_i) \sim N(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1, 2, 3$, where the $\boldsymbol{\mu}_i$'s and Σ_i's are unknown. Use the training set to compute the MLE of the $\boldsymbol{\mu}_i$'s and the Σ_i's. What is the error rate on the test set when the Bayes decision rule using the estimated parameters is employed for classification?

   (c) Suppose the form of the distributions of $p([x_1, x_2]^t|\omega_i)$, $i = 1, 2, 3$ is unknown. Assume that the training dataset can be used to estimate the density at a point using the Parzen window technique (a spherical Gaussian kernel with $h = 1$). What is the error rate on the test set when the Bayes decision rule is employed for classification?

   (d) Suppose the 1-nearest neighbor method is used for classifying the patterns in the test set. What is the error rate of this approach?

   (e) Repeat the above three classification procedures by varying the size of the training set as follows: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of each class. Plot the error rate as a function of the size of the training set for each of the 3 cases.

   (f) Discuss you results along the following lines:

      i. Which classifier results in the best performance on the test set?

      ii. Does the performance of a classifier change significantly depending upon the patterns used in the training set? Substantiate your answer with an actual experiment on this dataset.

iii. How does the performance of a classifier change as a function of the *number* of data patterns used to estimate its parameters?

iv. Do you think it is necessary for the number of training patterns per class to be the same?

2. [10 points] In a New York Times article from 2003 (see http://writing.upenn.edu/~afilreis/88v/kurzweil.html), Ray Kurzweil is quoted as saying "The real power of human thinking is based on recognizing patterns. The better computers get at pattern recognition, the more humanlike they will become." What are your thoughts on this statement?
