CS670 Artificial Intelligence

Lab #3: Multivariate Methods & Dimensionality Reduction

Out – 03/28/2011, Due – 04/08/2011

Programming homework

In this homework, you are given *Wine* dataset from *UCI Machine Learning Repository*. The data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The data format is given as *"[cultivar type], [feature #1], … , [feature #13]"*. Hence, the cultivar type of each vector is a class, and each data is a 13-dimensional vector. More descriptions for the wine dataset can be found in "http://archive.ics.uci.edu/ml/datasets/Wine".

(1) You should implement principal component analysis (PCA). ***pca = function(ndim)*** is the function you should write the PCA code on. The function input ***ndim*** represents the number of dimensions after the projection. For example, if ***ndim*** = 2, ***pca*** should return a [#data]-by-2 matrix, where each row is a projected vector of each instance. Do NOT use the built-in function in R that does PCA for you. **(25 points)**

(2) You should implement linear discriminant analysis (LDA). ***lda = function(ndim)*** is the function you should write the LDA code on. The function input ***ndim*** represents the number of dimensions after the projection. For example, if ***ndim*** = 2, ***lda*** should return a [#data]-by-2 matrix, where each row is a projected vector of each instance. **(25 points)**

(3) Write a function ***graph = function(proj, class)*** to show the results of dimensionality reduction when the projection space is 2-dimension. Thus, ***proj*** is the [#data]-by-2 matrix representing data projected onto 2-dimensional space, and ***class*** is the [#data]-dimensional vector representing class of each data. **(15 points)**

Written homework

Please write down the answers to a text file and submit the file with the above programming homework.

(4) What is the maximum dimension of the projection space for the LDA? And why?. **(15 points)**

(5) (Exercise 5.10 – problem #5) Let us say we have two variables x_1 and x_2, and we want to

make a quadratic fit using them, namely

$$f(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 (x_1)^2 + w_5 (x_2)^2.$$

How can we find w_i, $i = 0, \ldots, 5$, given a sample of $X = \{x_1^t, x_2^t, r^t\}$? You have to show your intermediate calculations. **(20 points)**

Submission Format

- Submit the [**#ID]_[FIRSTNAME]_[LASTNAME]_lab3.zip** file that contains below two files.
- For example, 20110123_John_Doe_lab3.zip

  (1) R file: **[#ID]_[FIRSTNAME]_[LASTNAME]_lab3.R**

  (2) docx or pdf file: **[#ID]_[FIRSTNAME]_[LASTNAME]_lab3.docx** (or .pdf)