BM 59D Machine Learning

Homework 3

Due: November 21, 2017

You are given two different sets of data, namely, $\mathcal{X} = \{\mathbf{x}^t, r^t\}_{t=1}^{200}$ and $\mathcal{Y} = \{\mathbf{y}^t, r^t\}_{t=1}^{200}$, where $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^3$, and $r^t = 0$ if the instance belongs to C_0, and $r^t = 1$ if it belongs to C_1. (You have 100 instances for each class, use the first 60 instances for training and the remaining 40 for validation.)

Perform each one of the following three options on both $\mathcal{X}$ and $\mathcal{Y}$, and report the associated confusion matrices. Compare the scores, and comment on the differences between the methods and between the structures of the datasets (the way they lie in space).

1) Adopting Bayesian decision theory, assuming Gaussian likelihoods for the classes, and using maximum likelihood (ML) estimators for the class parameters, train the classifier using the training set and test it on the validation set. Try all possible options for the covariance matrices, i.e.:

   a) Use distinct covariance matrices (ML estimates calculated separately for the two classes).

   b) Use a common covariance matrix for the two classes.

   c) (In addition to (b)) Assume a diagonal covariance matrix.

   d) (In addition to (c)) Assume equal variances.

2) Perform PCA. Let the new space be one dimensional. Report the proportion of variance explained associated with each set. After PCA, perform Bayesian decision with Gaussian likelihoods.

3) Perform LDA. After LDA, perform Bayesian decision with Gaussian likelihoods.

Please use MATLAB. You may find the built-in functions plot3(), repmat() and diag() useful. Please do not use the built-in or community-shared functions to calculate the probabilities, to predict the labels, or to perform PCA and LDA.

In your report, I prefer that you put all the confusion matrices on one page. Proper visualizations of the data before and after dimensionality reduction are appreciated.

Please send your report by e-mail (To: ipekbounee@gmail.com, Sbj: BM59D HW3), and give your codes in the appendix at the end of your report so that I can trace back if unexpected results appear. I prefer one single document including all the material rather than a folder including a report together with several m-files.