CS5691 : Pattern Recognition and Machine Learning

Programming Assignment 1 **Deadline: 12^{th} March 2020, 23:55 hrs**

Instructions:

1. You have to turn in the well-documented code along with a detailed report of the results of the experiments.
2. Any kind of plagiarism will be dealt with severely. Acknowledge any and every resource used.
3. Plot your data and analyze before proceeding.
4. Be precise with your explanations. Avoid verbosity. Put only relevant and best/worst results plot. Report size should be **10-12 pages** (single column, 11pt). You will get a heavy penalty if you make a longer report than this.
5. You can use any language for this assignment. Using Python or MATLAB would be easier.
6. Create a folder named "TeamNumber_TeamMember1RollNo_TeamMember2RollNo' (for e.g. "1_CS19S016_CS19S011). In this folder, you should have your report and a sub-folder "codes" which should have all your codes. Upload this folder(.zip) on Moodle. Please follow the naming convention strictly.
7. Please make only one submission for the team. No emailed reports will be accepted.

Bayesian Classifier

In this assignment, you are supposed to build the Bayesian classifiers for the datasets assigned to your team. The purpose of this assignment is the analysis of classification techniques and getting used to handling data in machine learning. Dataset for each team can be found here. It also has the sample plots required. Divide the points randomly for training(80%)and testing(20%).

Algorithm 1: Bayes Classifier

1. Compute prior probabilities for each class from the dataset.
2. Estimate class conditional densities using the maximum likelihood estimation.
3. Use Bayes rule to estimate the posterior probability $q_i(X), i = 0, 1, ...M-1$.
4. The Bayes classifier, h_B for the M-class case is:

$$h_B(X) = \alpha_i \quad \text{if} \quad \sum_{j=0}^{M-1} L(\alpha_i, C_j)\, q_j(X) \leq \sum_{j=0}^{M-1} L(\alpha_k, C_j)\, q_j(X), \forall k \quad \text{(break ties arbitrarily)}$$

where, $C_0, C_1, \cdots, C_{M-1}$ are the class labels, $L(\alpha_j, C_k)$ is the loss when classifier says α_j and true class is C_k, and $q_i(X)$ is the posterior probability.

---

1. Consider the following loss function L : $\begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 3 \\ 1 & 3 & 0 \end{bmatrix}$ and build a Bayesian model on datasets 1 and 2 for the following cases (You can refer Section 2.6 of "Pattern classification" book by [Duda et al. 2001] for theory): (4x5=20 marks)

   (a) Model 1: Naive Bayes classifier with covariance = $\mathcal{I}$.

   (b) Model 2: Naive Bayes classifier with covariance same for all classes.

   (c) Model 3: Naive Bayes classifier with covariance different for all classes.

   (d) Model 4: Bayes classifier with covariance same for all classes.

   (e) Model 5: Bayes classifier with covariance different for all classes.

   The report should include the following for both datasets (You can refer to sample plots here and Figures 2.6 and 2.9 of "Pattern classification" book by [Duda et al. 2001]):

   (i) Table of classification accuracies for all the models on training data and validation data. (1 marks)

   (ii) Confusion matrix on test set for the best model among those in parts (a)-(e). (1 marks)

   (iii) Decision boundary and decision surface of the best model. Superpose the training data on this plot. (1 marks)

   (iv) Contour curves and eigenvectors of the covariance matrix for the best model. (1 marks)

2. Consider Bayesian estimation of mean on one-dimensional Gaussian dataset 3. Suppose prior of the mean is $\mathbf{P}(\mu) \sim \mathcal{N}(\mu_0, \sigma_0)$. (4 marks)

   (a) Estimate σ by assuming that $\mu_0 = -1$. (1 marks)

   (b) Plot your estimated densities $\mathbf{P}(x/\mathcal{D})$ for $n = 10, 100, 1000$ with the following rates for $\sigma^2/\sigma_0^2 = 0.1, 1, 10, 100$. (3 marks)

Maximum Likelihood Estimation

3. Consider the classification of Gaussian data employing shrinkage of covariance matrices to a common one. Generate 20 training points from each of three equally probable three-dimensional Gaussian distributions $\mathcal{N}(\mu_i, \Sigma_i)$ with the following parameters: (8 marks)

$$\mu_1 = (0,0,0)^t, \Sigma_1 = diag[3,5,2]$$
$$\mu_2 = (1,5,-3)^t, \Sigma_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 6 \end{pmatrix}$$
$$\mu_3 = (0,0,0)^t, \Sigma_3 = 10\mathbb{I}$$

   (a) Report the estimated means and covariances of your data. (3 marks)

   (b) Shrink these estimated covariance matrices according to this equation (see Section 3.7.3 of "Pattern classification" book by [Duda et al. 2001] for theory)

$$\Sigma_i(\alpha) = \frac{(1-\alpha)n_i\Sigma_i + \alpha n\Sigma}{(1-\alpha)n_i + \alpha n}, \text{ for } 0 < \alpha < 1$$

   Plot the training error as a function of α, where $0 < \alpha < 1$. (3 marks)

   (c) Use your code from part (a) to generate 50 test points from each category. Plot the test error as a function of α. (2 marks)

Singular Value Decomposition (SVD)

4. Create a 100×100 matrix of random numbers between 0 and 1 such that each entry is highly correlated with the adjacent entries. Perform SVD on it and answer the following questions: (8 marks)

   (a) Compute the Frobenius norm of the matrix A. (1 marks)

   (b) What fraction of the Frobenius norm of A is captured by the top 10 singular vectors? (1 marks)

(c) What fraction of the Frobenius norm of A is captured by random 10 singular vectors? (1 marks)

(d) Plot a graph showing the number of singular vectors required to capture 50%, 75% & 95% of the data respectively? (Using the Frobenius norm) (1 marks)

(e) Write down your observations. (1 marks)

(f) Repeat all the above five parts for 100×100 matrix of statistically independent numbers between 0 and 1. (3 marks)

Principal Component Analysis (PCA)

5. Run the PCA algorithm on images here. The report should include the following for both experiments: (10 marks)

   (a) First convert it to grayscale. (2 marks)

   (b) Reconstruct the image using top N (N = 10%, 25%, 50%) principal components. Plot the reconstructed images along with their corresponding error image. How good is the quality of the reconstructed image?
   (*Hint: Use Frobenius norm to measure the quality of the reconstructed image.*) (3 marks)

   (c) Try random N (N = 10%) principal components instead of top N. Plot the reconstructed image along with its error image. Explain the observed trend in the result. (3 marks)

   (d) Plot a comparative graph of the reconstruction error vs N. (1 marks)

   (e) Write your inference in one or two lines. (1 marks)

Regression and Bias Variance

Download the required data from here.

6. Generate 100 sample points in the domain (0,1). For each sample point, generate a target value using the function assigned to your team, with additive Gaussian Noise having mean zero and variance 0.2. Divide the data set as 80:20 (train:test). Perform polynomial regression. In particular, do the following: (*Plot the data, the target function and the regression output. Also tabulate the coefficients obtained for each model.*) (10 marks)

   (a) Choose 10 points from the training data-set and perform regression for degrees: {1,3,6,9}. (4 marks)

(b) For each of the models above, analyze over-fitting by varying the data-set size. (2 marks)

(c) Show the scatter plot with target output t_n on x-axis and model output $y(x_n, w)$ on y-axis for the best performing model, for training data and test data. (2 marks)

(d) Plot the root-mean-square(RMS) error. (*See 1.3 in Bishop's book for definition.*)(2 marks)

For sample plots, see Figures 1.4, 1.5, 1.6 and Table 1.1 from "Pattern Recognition and Machine Learning" book by [Bishop et al. 2006]