Assignment # 1

SYDE 675 Winter 2020

The assignment can be done in Python or Matlab.

You need to submit both your report and the source code implementation for all questions. The report must be a single pdf and the source code must be a single .py or .m file. Please include brief comments in your code. Be sure to label all figures and include a legend where appropriate.

The due date for this assignment is **Feb 20st, 2020**. Please also note that late submission will be subject to a penalty of 20% deduction of the assignment mark per day.

---

1) Consider two classes described by the covariance matrices below (assume zero mean) (15)

$$\text{A. } \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad B.\ \Sigma = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix}$$

   a) For each matrix generate 1000 data samples and plot them on separate figures.
   b) For each case calculate first standard deviation contour as a function of the mean, eigenvalues, and eigenvectors. Show your calculation (Hint: consider distribution whitening from the tutorial). You may use preexisting functions for Eigen computation. Plot each contour on the respective plots from part (a).
   c) Calculate sample covariance matrices for each class using the data generated in part. Do not use a Python/Matlab function for computing the covariance.
   d) Explain the difference between the given covariance matrix for each class and the corresponding sample covariance matrix generated in (b). In which condition they can be the same?

2) Consider a 2D problem with 3 classes where each class is described by the following priors, mean vectors, and covariance matrices. (25)

$P(C_1) = 0.2$ $\qquad$ $P(C_2) = 0.7$ $\qquad$ $P(C_3) = 0.1$

$\mu_1 = [3 \;\; 2\,]^T$ $\qquad$ $\mu_2 = [5 \quad 4\,]^T$ $\qquad$ $\mu_3 = [2\;5\,]^T$

$$\Sigma_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 3 \end{bmatrix}$$

a) Create a program to plot the decision boundaries for a ML and MAP classifier. Plot the means and first standard deviation contours for each class. Discuss the differences between the decision boundaries.
b) Generate a 3000 sample dataset using the prior probabilities of each class. For both the ML and MAP classifiers: classify the generated dataset, calculate a confusion matrix, and calculate the experimental P(ε). Discuss the results.

3) The MNIST dataset contains a set of images containing the digits 0 to 9. Each image in the data set is a 28x28 image. The data is divided into two sets of images: a training set and a testing set. The MNIST dataset can be downloaded from http://yann.lecun.com/exdb/mnist/. Use only the training set to perform this part. (25)

a) Program PCA that takes X(DxN) and returns Y(dxN) where N is the number of samples, D is the number of input features, and d is the number of features selected by the PCA algorithm. Note that you must compute the PCA computation method by yourself. You may use preexisting functions for Eigen computation.
b) Propose a suitable d using proportion of variance (POV) =95%.
c) Program PCA reconstruction that takes Y_{PCA}(dxN) and returns $\hat{X}$ (DxN) (i.e., a reconstructed image). For different values of d= {1, 20, 40, 60, 80, …, 760, 784} reconstruct all samples and calculate the average mean square error (MSE). Plot MSE (y-axis) versus d (x-axis). Discuss the results.
d) Reconstruct a sample from the class of number '8' and show it as a 'png' image for d= {1, 10, 50, 250, 784}. Discuss the results.
e) For the values of d= {1, 2, 3, 4, …, 784} plot eigenvalues (y-axis) versus d (x-axis). Discuss the results.

4) Consider the attached file dataset3.txt. The first two columns of the data file show the feature of each sample and the last column illustrates its corresponding binary level. (25)

a) What is the cost function in logistic regression?
b) Estimate the parameters using stochastic gradient descent (SGD) method. You need to implement the SGD function for the optimization.
c) Plot the cost function along the epochs of the SGD.
d) Use the learned model to classify all training samples and report the accuracy.
e) Plot the data and show the class of the sample using different colors.
f) Plot the Decision boundary of the classifier.

5) Naïve Bayes Discussions: (10)
   a) What is the difference between Bayes and Naïve Bayes classifiers?
   b) In which situation, Naïve Bayes is equivalent to Bayes?
   c) In practice, Bayes classifier is not tractable in many applications. Explain, when Bayesian classifier can be practically used?
   d) Discuss why Bayes classifier is not tractable while Naïve Bayes is tractable.