**Classification Techniques**

**Problem one:** Design and implement general Multivariate Gaussian classifier for multiclass classification using python

1. Compare your implementation with the Gaussian implemented in sklearn library “*class sklearn.naive\_bayes.GaussianNB( )*” using the accuracy measure
2. Draw the training error and validation error curves on one graph for both algorithms as follows
3. First split your data into training set 70% and test set 30%
4. Use cross validation technique on train set to draw the training and validation error as follows

* Repeat the experiment 10 times, in first experiment use 10% of the training set for training and validation, in the second experiment used 20% of training set for training and validation and so on until the tenth experiment were use 100% of training set for training and validation.
* In each experiment divide the used set ***(i.e. the 10% or 20% .. or 100% of training set)*** into 10 folds where one-fold used for validation and the other 9-folds used for training, in which it will repeated 10 times each time record the training error and validation error and finally take the average for the training and validation
* After finishing the 10 experiments you will have 10 average values for training error and 10 average values for validation error
* Finally draw the training error curve and validation error curve with respect to training set size used in each experiment
* Show if overfitting or underfitting are appear in your algorithm if so, solve it and show the result after you solve it

1. Compare the training error, validation error with test error for both algorithms

**Problem two:** Design and implement general quadratic discriminant analysis (QDA) classifier for multiclass classification using python

1. Compare your implementation of the second algorithm (QDA) with the SVM implemented in sklearn library “*class sklearn.svm.SVC* ( )” using the accuracy measured
2. Draw the training error and validation error on one graph for both algorithms as follows
3. First split your data into training set 70% and test set 30%
4. Use cross validation technique on train set to draw the training and validation error as follows

* Repeat the experiment 10 times, in first experiment use 10% of the training set for training and validation, in the second experiment used 20% of training set for training and validation and so on until the tenth experiment were use 100% of training set for training and validation.
* In each experiment divide the used set into 10 folds where one-fold used for validation and the other 9-folds used for training, in which it will repeated 10 times each time record the training error and validation error and finally take the average for the training and validation
* After finishing the 10 experiments you will have 10 average values for training error and 10 average values for validation error
* Finally draw the training error curve and validation error curve with respect to training set size used in each experiment
* Show if overfitting or underfitting are appear in your algorithm if so, solve it and show the result after you solve it

1. Compare the training error, validation error with test error for both algorithms

Interpret your result on both problems and show which algorithm of the fourth algorithms give better result and explain why

**Note: verify your two problems on the data set given with assignment**