Intelligent Data Analysis – Spring 2018

Homework #2

**Due Date: Feb 16th, 2018, 9PM**

* Consider the two data files attached to this homework assignment. This data is taken from the UCI Machine Learning Repository and contains biomechanical features of some patients. The task is to predict whether the patient is normal or abnormal (call it Data2). The second dataset splits the abnormal group into two different diagnoses (call it Data3). Both datasets have same 310 feature vectors, six features, and a class column. Perform the following tasks with these datasets and submit the answers asked for in each task listed below.
* All your answers must be contained in a single pdf file. Upload this pdf file on Blackboard in response to the homework assignment.
* Use **svmtrain** function of Matlab or any other tool box that you are familiar with. If you are not using Matlab then include in your submission the name of the toolbox that you use and an example of commands issued for generating the classifier models. Do “help svmtrain” to learn the details about this matlab function (<https://www.mathworks.com/help/stats/svmtrain.html)>. Use all defaults for the parameter values except for the kernel function which should be chosen as stated in the questions.
* Each submission must be individual work of each student. Any plagiarism detected will be severely punished.

1. (25) Take Data2 and split it into a Training Set of 210 randomly selected data points. Use the remaining 100 data points as the Testing Set. Train a linear support vector machine (kernel\_function = linear) to build a classifier model. Use this model to predict the classes for the data in the Testing Set. Report the following:
   1. List all the support vectors of the data set. Examine the number of support vectors for each class and their attribute values and comment on what can be inferred from these values.
   2. Run the training and testing five times, each time selecting a different randomly selected set of training instances. Create the confusion matrix for this classifier using average number of true positives, false positives etc.
   3. Compute the accuracy, precision and recall values.
2. (25) Repeat Q#1 above with the only difference that this time train a non-linear SVM using rbf (radial basis function) for kernerl function.

|  |  |  |  |
| --- | --- | --- | --- |
| X | Y | Z | Class |
| 5 | 2 | 8 | 1 |
| -2 | 9 | 3 | 0 |
| 0 | 10 | 4 | 0 |
| 6 | 1 | 4 | 1 |
| 8 | -3 | 9 | 1 |

1. (25) Compare the performance values obtained in 1c and 2c above with the ones you received for the same data in Problem 1(c) of Homework#1. Comment on the differences and you observe and their possible causes/consequences.
2. (25) Consider a dataset of 3-D data points as shown here.
   1. Show every step of the working of the Perceptron Training Algorithm with this data **for only the first two epochs.**  Use (1, 2, 3, 4) as the initial weight vector.
   2. What is the error value, J(w), after each epoch?
3. (5) These 5 bonus points are for good organization and presentation of results in your answers.