MEEM 5990 Applied Machine Learning

Spring 2021

Project 4

Total Points: 50

**Support Vector Machine**

The objective of this part is to use Hard margin, Soft margin and kernel SVM on Sonar data set available at UCI Machine Learning Dataset Repository.

**Dataset information**

Information for this dataset is available at <http://archive.ics.uci.edu/ml/datasets/connectionist+bench+(sonar,+mines+vs.+rocks)>

**Deliverables:**

* A single code file which should run on a click when the sonar.mat file is placed in the working directory. (Functions are to be written at the end of file in MATLAB)
* A pdf file of all the results

**Following tasks are to be performed:**

1. Loading the data in MATLAB

load('sonar.mat')

After loading the data, you will have 2 variables in your workspace: X and Y. X is the data matrix where rows are observations. Y contains the class labels, either -1 or +1. There are 208 observations in this dataset.

Select random 25% of the data for testing. Remember to select respective data labels from Y matrix. The remaining data is training data.

You should have following matrices after this step,

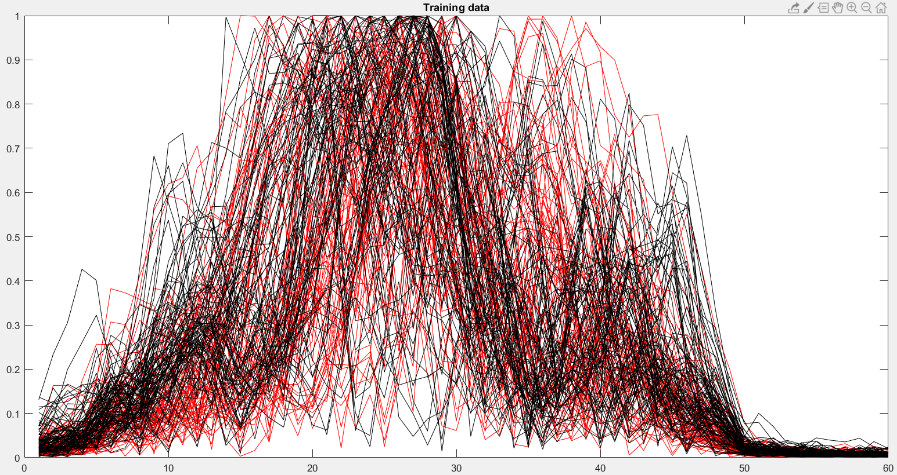
trainfeatures (156\*60)

trainlabels (156\*1)

testfeatures (52\*60)

testlabels(52\*1)

1. Plot training data and test data in **2 different plots**. Use different colors for different colors. Here is my training data with red color for label==-1 and black color for label==+1. Don’t forget to add titles to the plot. Legend to the plot can be in the plot or can be added as a caption.



1. Apply a hard margin SVM and report the testing accuracy. You can use inbuilt function for this, or you can code it on your own.
2. Apply a soft margin SVM and report the testing accuracy. Value of C should be used from previous project. Whatever you think was the best.

Till this point Project 4 is pretty much same as Project 3.

1. Apply kernel SVM with ‘rbf’ kernel using the inbuilt MATLAB option and report your testing accuracy. Provide transformation function for ‘rbf’ kernel.
2. Create a function ‘myGaussian’and use this function for kernelization.

kernelSVMModel = fitcsvm(train\_features,train\_labels,'KernelFunction', 'myGaussian');

‘rbf’ kernel has a hyper parameter . Find the optimum value of by performing 5-fold cross validation. Report 5 testing accuracies on 5 different sets. Report means testing accuracy. Use the optimum and report your testing accuracy.

1. Apply kernel SVM with any kernel of your choice other than ‘rbf’. Provide transformation function for the kernel.
2. Aggregate all results in the following table.

|  |  |
| --- | --- |
| **Method** | **Testing Accuracy** |
| Hard Margin SVM |  |
| Soft Margin SVM with C = …  (C value obtained using 5-fold CV) |  |
| Kernel SVM using the MATLAB inbuilt ‘rbf’ option |  |
| Kernel SVM using your own ‘myGaussian’ function at = …  ( value obtained using 5-fold CV) |  |
| Kernel SVM with …… Kernel (Task 7) |  |

**Point distribution:**

Task 2: 2 points

Task 3: 3 points

Task 4: 5 points

Task 5: 5 points

Task 6: 20 points

Task 7: 5 points

Table: 5 points

Readability of the code: 5 points

**Acknowledgements:**

The sonar data was obtained from UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.