

Report

Name: Dhruv Sangvikar
Net ID: dgs160230

Support Vector Machines

By definition of SVM, libSVM will generate hyperplanes using the given supervised training data with labels, and using the model try to apply and predict labels on the validation data.

The task was pretty straight forward. The libSVM package provided the makeFile to build the following executables:

- svm-scale
- svm-train
- svm-predict

Using the following commands did the task of training the model:

```
./svm-train -t 0 promoters_data/training.new promoters_data/model/model-kernel-0
```

To make the prediction and record the accuracy, used the following command:

```
./svm-predict promoters_data/validation.new promoters_data/model/model  
promoters_data/output
```

As data provided was already in the libsvm compatible format and range of values was not large, no preprocessing or scaling of data was required.

According to different kernels, the model was able to classify the validation data with different accuracy. Following table lists the kernel type and accuracy obtained on the validation data set.

Sr. No.	Kernel Type	LibSVM FlagValue (-t)	Accuracy	Correct Classification Count
1	linear	0	85.7143%	30/35
2	polynomial	1	74.2857%	26/35
3	radial basis	2	77.1429%	27/35
4	sigmoid	3	45.7143%	16/35

Observation:

It can be observed that the data labels are binary (labels 0,1). Also, the number of samples in the training data is comparable to the number of features. Both of these factors account for the high accuracy shown using linear kernel which normally works better on such kind of data. The few number of training samples accounts for the sigmoid function working so poorly. The radial basis kernel, which is considered to be a good default function, works better than sigmoid or polynomial.