CS 312 – Artificial Intelligence Laboratory

Lab 6 Report

180010020 – Karlapati Sree Ramu

180010034 – Singamaneni Meghana

Problem Description:

We will use SVM algorithm to classify emails into SPAM/NOT SPAM categories.

We observe the efficiency of the algorithm by using different kernel functions and changing generalisation constant C.

We use below kernel functions with different generalisation constant values:

1) Linear kernel function
2) Quadratic kernel function
3) RBF kernel function

Libraries used:

We used “svm” library from “sklearn” package in python.

from sklearn import svm

clf = svm.SVC(kernel = ‘linear’, C = c) # for Linear kernel function

clf = svm.SVC(kernel = ‘poly’, degree = 2, C = c) # for Quadratic kernel function

clf = svm.SVC(kernel = ‘RBF’, C = c) # for RBF kernel function

clf.fit(training_data, training_labels) # Classifying training data

Methodology:

Linear kernel function:

Linear kernel function is used when data is linearly separable. In this method, data can be separated by a single line. If there are more no. of features, linear kernel-based separation gives better results. Linear kernel-based classification gives better results when no pre-processing happened.

Accuracy: Around 93% (Without pre-processing for the data)

Accuracy: Around 88% (If data is normalized)

SVM Package used: svm.SVC and kernel function is linear

(svm.SVC(kernel = 'linear', C = c))

Quadratic kernel function:

Here, instead of a line we use a Quadratic kernel function to classify the given data.

Quadratic kernel function is the most used polynomial kernel because higher degree may cause over-fitting and leads to mis-classification.

Here, we got much better results if we normalize the given data.

Accuracy: Around 68% (Without pre-processing for the data)

Accuracy: Around 88% (If data is normalized)

SVM Package used: svm.SVC and kernel function is poly of degree 2

(svm.SVC(kernel = 'poly', degree = 2, C = c))

RBF kernel function:

RBF means Radial basis function. It is the most popular kernel function commonly used in SVM classification. Here, we got much better results if we normalize the given data.

By using this kernel function, we got much better result in less time

Accuracy: Around 70% (Without pre-processing for the data)

Accuracy: Around 93% (If data is normalized)

SVM Package used: svm.SVC and kernel function is RBF

(svm.SVC(kernel = 'RBF', C = c))

Experimental results:

Kernel function used	Accuracy	Execution time
Linear kernel function	93.07%	103.69 seconds
Quadratic kernel function	88.5%	0.42857 seconds
RBF kernel function	93%	0.505 seconds

Although Linear kernel function is giving more accuracy but it is taking more time for executing and giving the output line.

Quadratic function is taking very less time for execution but accuracy of the results is some what less compared to Linear kernel function.

RBF kernel function is also taking less time for execution and accuracy of the results is also more.

By taking both computational time and accuracy of the algorithm, I prefer to use RBF kernel function for SVM classification.

Varying Generalization constant C:

Table for Linear kernel function:

Generalization constant	Accuracy	Execution time
0.01	91.79%	2.985 seconds
0.05	93.005%	28.14 seconds
0.1	92.8414%	37.813 seconds
0.3	92.72%	174.72 seconds
0.5	92.5%	247.6546 seconds
1	92.648%	135.7272 seconds
2	92%	410.42786 seconds
5	92.1458%	909.388 seconds
10	91.791577%	937.2966 seconds

From the above table, we can observe that execution time for Linear kernel function is increasing by increasing the value of the Generalization constant.

Accuracy of the Linear kernel function is decreasing by increasing or decreasing Generalization constant value a lot.

We can say that accuracy will be high at Generalization constant = 0.05 to 0.1 and execution time is not high. So, we can say that 0.05 to 1 is the optimal value of Generalization constant for Linear kernel function.

Table for Quadratic kernel function:

Generalization constant	Accuracy	Execution time
0.01	75.66%	0.837645 seconds
0.05	82.44%	0.6628778 seconds
0.1	82.7266%	0.589 seconds
0.3	86.36688%	0.5127375 seconds
0.7	88.7937%	0.46455 seconds
1	89.5%	0.425 seconds
2	89.86439%	0.39453 seconds
3	90.0785%	0.385333 seconds
5	90.935%	0.372962 seconds
10	90.935%	0.44737 seconds

From the above table, we can observe that execution time for Quadratic kernel function is decreasing by increasing the value of the Generalization constant to some extent. At higher values of Generalization constant, Quadratic kernel function is taking more time for execution

Accuracy of the Quadratic kernel function is increasing by increasing the value of Generalization constant.

We can say that accuracy will be high at Generalization constant = 5 and execution time is also very less. So, 5 is the optimal value for Generalization constant.

Table for RBF kernel function:

Generalization constant	Accuracy	Execution time
0.01	80.728%	1.2057 seconds
0.05	88.294%	0.8723724 seconds
0.1	88.72234%	0.75 seconds
0.3	91.3633%	0.61 seconds
0.7	92.29122%	0.535 seconds
1	93.576%	0.507 seconds
3	92.9336%	0.45 seconds
5	92.8622%	0.4404 seconds
10	93.79%	0.4844 seconds

From the above table, we can observe that execution time for RBF kernel function is decreasing by increasing the value of the Generalization constant to some extent. At higher values of Generalization constant, RBF kernel function is taking more time for execution

Accuracy of the RBF kernel function is increasing by increasing the value of Generalization constant to some extent.

We can say that accuracy will be high at Generalization constant = 10 and execution time is also very less. So, 10 is the optimal value for Generalization constant.

Finally, I prefer RBF kernel function for SVM classification. By using this kernel function we are getting most accurate results in less time.