

CS 312: Artificial Intelligence Laboratory

Task 6: Machine Learning

Support Vector Machine Classifier

Problem Statement:

Spam email classification using Support Vector Machine: In this assignment you will use a SVM to classify emails into spam or non-spam categories. And report the classification accuracy for various SVM parameters and kernel functions.

Data Set Description:

An email is represented by various features like frequency of occurrences of certain keywords, length of capitalized words etc. A data set containing about 4601 instances are available in this link (data folder): [Link](#)

The data format is also described in the above link. You have to randomly pick 70% of the data set as training data and the remaining as test data.

Assignment Tasks:

In this assignment you can use any SVM package to classify the above data set. You should use one of the following languages: c/C++/Java/Python. You have to study the performance of the SVM algorithms.

Submission:

Please submit a zip file `<Group_number>.zip` with the following contents

1. Program: `<group_number>.<extension>` (e.g., 1.c/cpp)
2. Report: `<group_number>.<extension>` (e.g., 1.pdf). Report should be in pdf format.
3. Readme file: readme.txt (Execution details)

Report Format :

The report should contain the following sections:

1. Mention library which you are using.
2. Methodology: Details of the SVM package used.
3. Experimental Results:
 - i. You have to use each of the following three kernel functions (a) Linear, (b) Quadratic, (c) RBF.
 - ii. For each of the kernels, you have to report training and test set classification accuracy for the best value of generalization constant C . The best C value is the one which provides the best test set accuracy that you have found out by trial of different values of C . Report accuracies in the form of a comparison table, along with the values of C .

Evaluation Criteria: Total:50 Marks

Code correctness: 25

Report: 20

Code Quality: 5

Deadline: 11:59 PM 16 March 2022

Late Submission Policy: *Penalty of 10% will be issued per day if the deadline is not met. If found copied, 0% score will be awarded*

For Reference :

[Support Vector Machines](#)