Support Vector Machine

Nisha Akole, G V V Sharma*

Contents

Abstract—**This manual provides a brief description on how to implement Support Vector Machine, a classifier technique.**

1. Objective

Our objective is perform classification of datasets and decide class for new data points.

2. Load Dataset

Download the dataset file from following link in the folder where you want to write code for SVM.

> https://github.com/prabhatrai111/Commensal−Radar

3. About SVM

Support Vector Machine(SVM) is supervised machine learning algorithm which can be used for both classification or regression challenges. It is well known for its kernels, to classify non-linear data. SVM known as a discriminative classifier as it separates data points using a hyperplane with large margin.

- Support Vector:
  These are the data points closest to hyperplane and helps to define a proper margin.

- Hyperplane:
  A plane which separates set of objects into different classes.

- Margin:
  A distance from hyperplane which separates classes. Larger margin is considered as good for classification and avoids overlapping.

SVM can be divided into two types based on dataset and kernel used to classify it

*The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

A. Simple SVM

In Simple SVM, parameters are linear. Kernel used is also a linear kernel to classify a linear data.

> link for linear kernel code

B. Kernel SVM

Kernel SVM is used specially to classify non-linear data by using different kernels such as polynomial, gaussian radial basis function(rbf) and sigmoid.

> link for 2 non−linear kernel code

4. Evaluating the Algorithm

Confusion matrix, accuracy, precision and recall are metrices used for checking the algorithm correctness. Scikit-learn is a free machine learning library for python, mostly used for various algorithms like support vector machine, principal component analysis, random forest etc. Sklearn contains classification_report and confusion_matrix for classification tasks.

5. Tuning Hyperparameters

- Kernel: SVM will give different output for different kernels. Use of kernel is totally dependent on the type of dataset. To get optimized output, one has to test different kernels for the good results.

- Regularization Parameter(C): It is used to calculate error and helps to reduce it. C is also used to decide margin. A small value of C will create a small margin around hyperplane and large value will create large margin. By this parameter, we cancontrols trade off between smooth decision boundary and error.

- Gamma: A lower value of gamma will underfit the data and higher value of gamma will overfit the data.

6. SVM Pros and Cons

A. Pros

- Works really well when no. of dimensions is greater than no. of samples i.e effective in high dimensional spaces.
- It gives good accuracy and uses less memory because they use a subset training points in the decision phase.
- SVM works well with clear margin of separation.
- Risk of overfitting is less in SVM.

B. Cons

- It takes more time for training. Hence, not suitable for large datasets.
- Works poorly with overlapping classes and also type of kernel used will also affect the output of SVM. Choosing good kernel is not easy.
- The SVM hyper-parameters are cost and gamma. It is hard to visualize their effect.