SVM and Kernels

Epoch IIT Hyderabad

Arin Aggarwal
MA22BTECH11006

1 Introduction

Svupport Vector Machine(SVM) is a powerful supervised model used for small but complex datasets generally used for classification problems.The aim is to find the best hyperplane that separates the 2 classes.

It can have 2 types:

1) **Linear SVM**: This works by simply selecting best possible hyperplane for data.
2) **Non-Linear SVM**: This requires use of kernels .

One of the most intuitive SVM models is the Maximum Margin Classifiers where the support vectors are used to make hyperplane giving maximum margin. This model fails when there are significant outliers and/or the data requires transformation.

2 Important Terms

1) **Support Vectors** :Support vectors are the data points that lie closest to the decision surface (or hyperplane). They are the data points most difficult to classify. They have direct bearing on the optimum location of the hyperplane.
2) **Hyperplane** : Hyperplane is the dividing boundary between classes. A data point on either side of the hyperplane is classiffied into 2 classes. Its dimension is one less than that of the space we are working in.
3) **Margin** :The margin is the distance between the decision boundary and the support vectors. An SVM aims to maximize this margin to improve generalization and reduce overfitting. It has 2 types:
   (i **Soft Margin** : When we want to allow some misclassifications in the hope of achieving better generality, we can opt for a soft margin for our classifier.
   (ii **Hard Margin** : When we don't want to have any misclassifications, we use SVM with a hard margin.

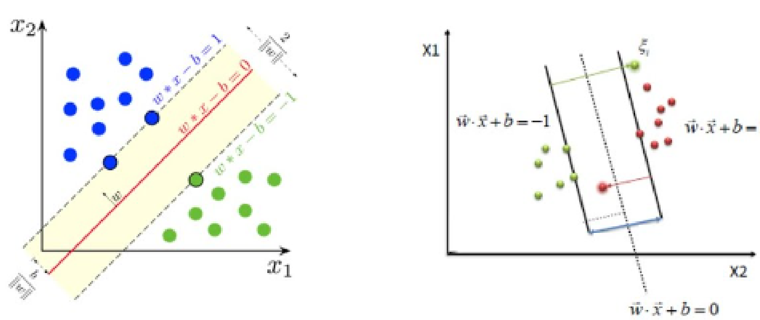
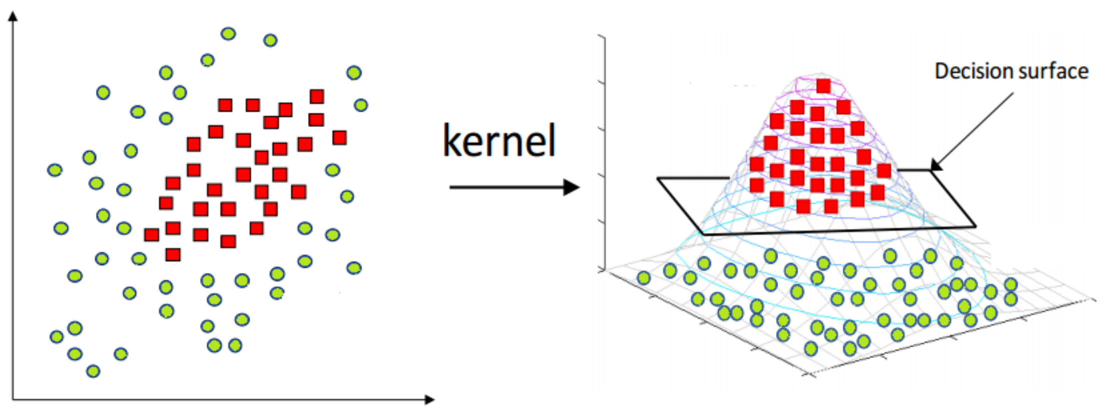


Fig. 3: Hard and Soft Margin respectively

Here we can see the bias-variance tradeoff at work.

4) **Kernels**: Kernel is the mathematical function, which is used in SVM to map the original input data points into high-dimensional feature spaces, so, that the hyperplane can be easily found out even if the data points are not linearly separable in the original input space. For this, Kernel trick is used which uses dot products for transformation to avoid complex computation.



3 Kernels

There are 2 main types of kernels:

1) **Polynomial Kernel**:

$$K(x_i, x_j) = (x_i.x_j + c)^d$$

where

$$d = degree\ of\ the\ kernel$$

$$x_i.x_j = dot\ product$$

2) **Radial Basis Function**:

$$K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}$$

sometimes parameterized using

$$\gamma = \frac{1}{2\sigma^2}$$

where

$$\sigma = variance\ and\ our\ hyperparameter$$

Mathematically, it is a dot product that has coordinates for an infite number of dimensions.

4 Fine-Tuning

We need to set optimal values for our parameters like d in polynomial kernel and γ in RBF kernel. This can be done using parameter like C which essentially decides the bias-variance tradeoff. A smaller C enables more misclassification, while a larger C imposes a stricter margin. Cross-Validation techniques are also used for tuning the parameters.