

Weekly Report 2 - SVM

Ganji Varshitha
AI20BTECH11009

Introduction

SVM is a supervised learning model used for both classification and regression. It creates decision boundary which divides the training data into classes, hence making it a discriminative classifier. It is non parametric as it does not take any information from the training data.

Algorithm

The algorithm creates a decision boundary in n-dimensional plane so as to classify the new points.

There are two types of SVM: linear SVM and non-linear SVM.

Linear SVM

Goal is to create a straight which separates the points of 2 classes. There are infinite lines which can be drawn.

Margin of the classifier is the width that the boundary could be increased by before hitting a data point.

In order to avoid misclassification we need to maximise the margin of the classifier, i.e we draw a line between the 2 closest points that keeps the other data points separated. Therefore, the data points that the margin pushes up against are called **Support Vectors**.

We can formulate the maximum margin problem as quadratic optimization problem.

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}^\top \mathbf{w} \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \end{aligned} \tag{1}$$

After computing optimal \mathbf{w} and b , the classifier is given by $\text{sign}(\mathbf{w}^\top \mathbf{x} + b)$. We assumed that all the positive points have $\mathbf{w}^\top \mathbf{x} + b > 1$ and negative points have $\mathbf{w}^\top \mathbf{x} + b < -1$ in the training dataset.

From the knowledge of Primal and duality, we can say that strong duality holds. The Lagrangian of the problem is

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n \alpha_i (1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)) \tag{2}$$

Since optimal solution for primal is same as that of dual, solving the dual with KKT conditions, we have

$$\begin{aligned} \max_{\alpha \geq 0} \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n \alpha_i (1 - y_i (\mathbf{w}^\top \mathbf{x}_i + b)) \\ \frac{\partial L}{\partial \mathbf{w}} = 0 \implies \mathbf{w} = \sum_j \alpha_j y_j \mathbf{x}_j \\ \frac{\partial L}{\partial b} = 0 \implies \sum_j \alpha_j y_j = 0 \end{aligned} \quad (3)$$

Substituting values of \mathbf{w}, b in the dual, we get

$$\begin{aligned} \max \sum_{k=1}^R \alpha_k - \frac{1}{2} \sum_{k=1}^R \sum_{l=1}^R \alpha_k \alpha_l Q_{kl} \text{ where } Q_{kl} = y_k y_l (x_k \cdot x_l) \\ \text{s.t } \alpha_k \geq 0 \quad \forall k \quad \sum_{k=1}^R \alpha_k y_k = 0 \end{aligned} \quad (4)$$

Solving the maximisation problem, we get optimal values of \mathbf{w}, b as

$$\mathbf{w} = \frac{1}{2} \sum_{k=1}^R \alpha_k y_k \mathbf{x}_k \quad (5)$$

$$b = -y_i (y_i (x_i \cdot \mathbf{w}) - 1) \quad (6)$$

Non Linear SVM

The best hyperplane is found by using the basis functions i.e kernel functions which classifies in the transformed high dimensional feature space.

The dot product in eq (4) is replaced by kernel functions.

Common kernels include polynomial and gaussian kernels.

Polynomial kernel: $k(\mathbf{x}_i \cdot \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$

Gaussian kernel: $k(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp \left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2\sigma^2} \right)$

Questions

1. Is the entire training set involved in predicting?

Solution:

No, only support vectors are used. No other points affect the decision boundary.

2. Which data points have non zero alpha weights in the dual problem?

A. All the training data points

B. Support vectors

Solution:

B. Support vectors.

3. Does SVM output a conditional probability of class?

Solution:

No.

4. Explain the methods to deal with multi class problem.

Solution:

There are two approaches: One vs One and One vs All.

One vs One : For n classes we will have $n(n-1)/2$ svm's for every pair of classes

One vs All: We have n svm's where datapoints belonging to one class are selected against rest of the data points.

5. Does changing the normalization constant in linear classifier affect the model?

Solution:

No.