

## Lecture 4c: Support Vector Machines (SVMs)

*Lecturer: Jeffrey Varner***Disclaimer:** *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

## 1 Introduction

In this lecture, we will discuss Support Vector Machines (SVMs). SVMs are a powerful class of supervised learning algorithms that can be used for classification and regression tasks. SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is a hyperplane that separates the data into classes. The goal of SVMs is to find the optimal decision plane that maximizes the margin between the classes. The margin is the distance between the decision plane and the closest data points from each class. SVMs are particularly useful for high-dimensional data and can handle non-linear decision boundaries using the kernel trick. In this lecture, we will discuss the basic concepts of SVMs, including the optimization problem, the kernel trick, and the soft margin SVM. We will also discuss the advantages and disadvantages of SVMs and provide an example of using SVMs for classification.

## 2 Basic Concepts of SVMs

Support Vector Machines (SVMs) are a class of supervised learning algorithms that can be used for classification and regression tasks. The goal of SVMs is to find the optimal decision plane that maximizes the margin between the classes. The decision plane is defined by the equation:

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (1)$$

where  $\mathbf{w}$  is the weight vector,  $\mathbf{x}$  is the data point, and  $b$  is the bias term. To estimate the decision plane, SVMs use the training data to find the optimal weight vector  $\mathbf{w}$  and bias term  $b$  that separate the data into classes. These parameters are estimated by solving an optimization problem that minimizes the norm of the weight vector subject to the constraints that the data points are correctly classified. The optimization problem for SVMs can be formulated as:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2)$$

subject to the constraints:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, n \quad (3)$$

where  $\mathbf{w}$  is the weight vector,  $b$  is the bias term,  $\mathbf{x}_i$  is the  $i$ -th data point,  $y_i$  is the label of the  $i$ -th data point, and  $n$  is the number of data points.

### 2.1 The Kernel Trick

The kernel trick is a technique used in SVMs to handle non-linear decision boundaries. The kernel trick allows SVMs to implicitly map the input data into a higher-dimensional space where the data is linearly separable. This is achieved by using a kernel function that computes the dot product of the data points

in the higher-dimensional space. The most common kernel functions used in SVMs are the linear kernel, polynomial kernel, and radial basis function (RBF) kernel. The linear kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (4)$$

The polynomial kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d \quad (5)$$

where  $\gamma$  is a scaling factor,  $r$  is a coefficient, and  $d$  is the degree of the polynomial. The RBF kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (6)$$

where  $\gamma$  is a scaling factor. Given a kernel function  $K(\mathbf{x}_i, \mathbf{x}_j)$ , the decision function of the SVM can be written as:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (7)$$

where  $\alpha_i$  are the Lagrange multipliers,  $y_i$  are the labels of the data points, and  $b$  is the bias term.

## 2.2 Soft Margin SVM

The soft margin SVM is an extension of the basic SVM that allows for some misclassification errors. The soft margin SVM introduces a slack variable  $\xi_i$  for each data point, which measures the distance of the data point from the decision plane. The optimization problem for the soft margin SVM can be formulated as follows:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (8)$$

subject to the constraints:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall i = 1, \dots, n \quad (9)$$

## 3 Summary and Conclusions

In this lecture, we discussed the basic concepts of Support Vector Machines (SVMs). SVMs are a class of supervised learning algorithms that can be used for classification and regression tasks.

## References