

Lecture 3a: The Perceptron for Binary Classification Problems

*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 introduce the perceptron algorithm for binary classification problems. The perceptron is a simple linear classifier that can be used to separate two classes of data points. The perceptron algorithm is a simple iterative algorithm that incrementally updates the weights of the linear classifier to minimize the classification error. We will first introduce the perceptron algorithm and then discuss its convergence properties, i.e., when the algorithm converges to a solution and when it does not. The key concepts covered in this lecture include:

- The perceptron algorithm for binary classification problems.
- The convergence properties of the perceptron algorithm.

2 The Perceptron Algorithm

The Perceptron (1) is a simple yet powerful algorithm used in machine learning for binary classification tasks. It operates by *incrementally* learning a linear decision boundary (linear regression model) that separates two classes based on input features by directly mapping the continuous output to a class such as $\sigma : \mathbb{R} \rightarrow \{-1, +1\}$, where the output function is $\sigma(\star) = \text{sign}(\star)$. Suppose there exists a data set $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ with N examples, with labels $y_i \in \{-1, 1\}$, and features $\mathbf{x}_i \in \mathbb{R}^n$. The Perceptron computes label y for feature vector \mathbf{x} as:

$$y = \text{sign}(\mathbf{w}^T \cdot \mathbf{x})$$

where $\mathbf{w} = (w_1, \dots, w_N, b)$ is a vector of weights w and a bias b , $\mathbf{x} = (x_1, \dots, x_n, 1)$ is a feature vector, and $\text{sign}(z)$ is the sign function:

$$\text{sign}(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

If data set \mathcal{D} is linearly separable, the Perceptron will find a separating hyperplane in a finite number of passes through \mathcal{D} . However, if the data set \mathcal{D} is not linearly separable, the Perceptron will not converge. Pusedo code for the perceptron algorithm is shown in Algorithm 1.

Algorithm 1 The Perceptron Algorithm

```

1: Input:  $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$ , tolerance  $\epsilon \geq 0$ , maximum iterations maxiter
2: Features:  $\mathbf{x}_i = (x_{i1}, \dots, x_{in}, 1)$  are augmented with a bias term, labels  $y_i \in \{-1, 1\}$ .
3: Output: Classifier parameters  $\beta = (w_1, \dots, w_n, b)$ 
4:  $\beta \leftarrow \text{rand}$  ▷ Initialize parameter vector  $\beta$  to a random vector
5:  $i \leftarrow 0$  ▷ Initialize the loop counter to zero
6: while true do ▷ Repeat until stopping criterion is met
7:   error  $\leftarrow 0$  ▷ Initialize the error count to zero for this pass through  $\mathcal{D}$ 
8:   for  $(\mathbf{x}, y) \in \mathcal{D}$  do ▷ Iterate over each pair  $(\mathbf{x}, y)$  in data set  $\mathcal{D}$ 
9:     if  $y \cdot (\mathbf{x}^T \cdot \beta) \leq 0$  then ▷ Oops! The data pair  $(\mathbf{x}, y)$  is misclassified
10:       $\beta \leftarrow \beta + y \cdot \mathbf{x}$  ▷ Update the weight vector  $\beta$ 
11:      error  $\leftarrow$  error + 1 ▷ Increment the error count
12:     end if
13:   end for
14:   if error  $\leq \epsilon$  or  $i \geq \text{maxiter}$  then ▷ Stopping criterion: tolerance or max iterations?
15:     break ▷ Exit the training loop
16:   end if
17:    $i \leftarrow i + 1$  ▷ Increment the loop counter and repeat
18: end while

```

3 Summary

In this lecture, we introduced the perceptron algorithm for binary classification problems. The perceptron is a simple linear classifier that can be used to separate two classes of data points. The perceptron algorithm is a simple iterative algorithm that incrementally updates the weights of the linear classifier to minimize the classification error. Thus, it is one of the first examples of an online learning algorithm, i.e., an algorithm that learns from data in an incremental fashion. The perceptron algorithm is guaranteed to converge to a solution if the data set is linearly separable. However, if the data set is not linearly separable, the perceptron algorithm will not converge to a perfect solution. If we are willing to accept some classification errors, we can use the perceptron algorithm to find a separating hyperplane in a finite number of passes through the data set, even if the data set is not linearly separable.

References

1. Rosenblatt F. Perceptron Simulation Experiments. Proceedings of the IRE. 1960;48(3):301–309. doi:10.1109/JRPROC.1960.287598.