Study Guide: Linear Predictors

Student Name

October 9, 2024

1 Introduction

Linear predictors are one of the most fundamental tools in machine learning. They form the basis for many learning algorithms, including classifiers and regression models. In this chapter, we explore linear predictors in the context of halfspaces, linear regression, and logistic regression.

2 Halfspaces (Section 9.1)

A halfspace is defined by a weight vector $w \in \mathbb{R}^d$ and a threshold $b \in \mathbb{R}$. A prediction for an input x is given by the sign of the linear function:

$$h(x) = \text{sign}(w^\top x + b).$$

The decision boundary is the hyperplane defined by $w^\top x + b = 0$.

2.1 Linear Programming for Halfspaces (Section 9.1.1)

We can use linear programming to find a separating hyperplane for linearly separable data. The goal is to minimize a loss function subject to the constraints that the predictions are correct for all training examples.

2.2 Perceptron Algorithm (Section 9.1.2)

The Perceptron algorithm iteratively adjusts the weight vector whenever a mistake is made. The update rule is:

$$w \leftarrow w + y_t x_t,$$

where (x_t, y_t) is a misclassified example.

2.3 VC Dimension of Halfspaces (Section 9.1.3)

The VC dimension of the class of halfspaces in $\mathbb{R}^d$ is $d + 1$. This means that this class can shatter at most $d + 1$ points.

3 Linear Regression (Section 9.2)

Linear regression aims to fit a linear model to data by minimizing the sum of squared errors.

3.1 Least Squares (Section 9.2.1)

Given a dataset $\{(x_i, y_i)\}_{i=1}^n$, the least squares solution minimizes the empirical loss:

$$L(w) = \frac{1}{n}\sum_{i=1}^{n}(w^\top x_i - y_i)^2.$$

The solution can be found analytically by solving the normal equations:

$$w = (X^\top X)^{-1}X^\top Y.$$

3.2 Polynomial Regression (Section 9.2.2)

Linear regression can be extended to handle nonlinear relationships by using polynomial features. For instance, we can map each input x to a higher-dimensional space containing polynomial terms.

4 Logistic Regression (Section 9.3)

Logistic regression is used for binary classification. The model outputs the probability that a given input belongs to a particular class, using the logistic function:

$$P(y = 1|x) = \frac{1}{1 + \exp(-w^\top x)}.$$

The weights are learned by minimizing the log-likelihood function:

$$L(w) = \sum_{i=1}^{n}\log(1 + \exp(-y_i w^\top x_i)).$$

5 Summary (Section 9.4)

- Linear predictors are foundational for both classification and regression tasks.
- Halfspaces, linear regression, and logistic regression provide flexible tools for a variety of learning problems.
- The VC dimension of linear classifiers gives insight into their complexity and generalization ability.