My Machine Learning Guide

Fawaz Shah

Contents

Part I.
Supervised Learning

Supervised learning methods can be split into two types: regression problems and classification problems.

In all supervised learning methods we have a **training set**, which is a collection of data which we already know the right answers for.

1. Regression

Regression methods are used for predicting continuous data.

1.1. Cost functions

A popular cost function is the **mean squared error**.

$$E = \frac{1}{2n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 \tag{1}$$

1.2. Linear regression

We try to fit a straight line to some data.
We have two variables, x and y.
Our line of best fit will be of the form:

$$f(p_0, p_1) = p_1 x + p_0 \tag{2}$$

where we must choose the optimum parameters (p_0, p_1) such that our cost function $E(p_0, p_1)$ is minimized.

This is achieved by gradient descent.

$$p_0 := p_0 - \alpha\frac{\partial}{\partial p_0}E(p_0, p_1) \tag{3}$$

$$p_1 := p_1 - \alpha\frac{\partial}{\partial p_1}E(p_0, p_1) \tag{4}$$

$$\tag{5}$$

The p_i will converge to their optimal value. α is what is called the **learning rate**, which decides how fast the parameters converge.

Deciding the optimal learning rate for the problem is tricky. If the learning rate is too slow then it will take a long time for the parameters to converge. If the learning rate is too fast then the parameters may overshoot the minimum point.

It is important that the parameters are updated *simultaneously* at each iteration.

1.3. Logistic regression

1 variable
More than 1 variable

2. Classification

Classification methods are used for predicting discrete data. We try to group data into categories.

2.1. PCA - Principal Component Analysis

Dimensionality reduction

Part II.
Unsupervised Learning

3. Clustering

3.1. K-means clustering

Part III.
Reinforcement Learning