Theory of linear regression

Chris Miles

January 11, 2021

1 Overview

We are interested in predicting some output value y from knowing a collection of input features $x_1, x_2, \ldots x_n$ represented as a vector $\mathbf{x} = (x_1, x_2, \ldots x_n)$. Suppose we have examples of the pair $(\mathbf{x}^{(k)}, y^{(k)})$ for $k = 1, 2, \ldots n$ that make up our dataset D from which we would like to infer the relationship between $\mathbf{x}$ and y. This is the task of many supervised learning problems. Linear regression makes a number of assumptions about the relationship between $\mathbf{x}$ and y. It models the relationship as

$$\hat{y}(\mathbf{x}; \mathbf{m}, b) = \mathbf{m} \cdot \mathbf{x} + b$$

The goal is to fit parameters $\mathbf{m}$ and b in the above equation to the dataset D such that the values for $y^{(k)}$ and the predicted $\hat{y}^{(k)} = \mathbf{m} \cdot \mathbf{x}^{(k)} + b$ are close on average across all k. More specifically, linear regression fits $\mathbf{m}$ and b through the following cost function:

$$J(\mathbf{m}, b) = \frac{1}{n} \sum_{k=1}^{n} |\hat{y}(\mathbf{x}^{(k)}; \mathbf{m}, b) - y^{(k)}|^2$$

which is the mean square error between the predicted $\hat{y}$ and the true value y for each point.