

LEAST SQUARES LINEAR PREDICTORS

Why

What is the best linear predictor if we choose according to a squared loss function.

Definition

Let $X \in \mathbb{R}^{n \times d}$ and $y \in \mathbb{R}^d$. In other words, we have a paired dataset of records with inputs in \mathbb{R}^d (the rows of X) and outputs in \mathbb{R} (the elements of y).

A least squares linear predictor or linear least squares predictor is a linear transformation $f: \mathbb{R}^d \to \mathbb{R}$ (the field is \mathbb{R}) which minimizes

$$\frac{1}{n} \sum_{i=1}^{n} (f(x^{i}) - y_{i})^{2}.$$

over the dataset of pairs $(x^1, y_1), \ldots, (x^n, y_n) \in \mathbf{R}^d \times \mathbf{R}$ where x^i is the *i*th row of X for $i = 1, \ldots, n$.

The set of linear functions from \mathbb{R}^d to \mathbb{R} is in one-to-one correspondence with \mathbb{R}^d . So we want to find $\theta \in \mathbb{R}^d$ to minimize

$$\frac{1}{n}||X\theta - y||^2.$$

Solution

Proposition 1. There exists a unique linear least squares predictor and its parameters are given by $(X^{\top}X)^{-1}X^{\top}y$.¹

¹Future editions will include an account.

