Why

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

Definition

Suppose we have a paired dataset of n records with inputs in $\mathbf{R}^d$ and outputs in $\mathbf{R}$. A *weighted least squares linear predictor* for nonnegative weights $w \in \mathbf{R}^n$, $w \geq 0$, is a linear transformation $f : \mathbf{R}^d \to \mathbf{R}$ (the field is $\mathbf{R}$) which minimizes

$$\frac{1}{n} \sum_{i=1}^{n} w_i (y_i - x^\top a^i)^2.$$

Some authors refer to this process of selecting a linear predictor as the *weighted least-squares problem*.

Define $W \in \mathbf{R}^{n \times n}$ so that $W_{ii} = w_i$ and $W_{ij} = 0$ when $i \neq j$. So, in particular, W is a diagonal matrix. We want to find x to minimize

$$\|W(Ax - y)\|$$

Solution

Proposition 1. *There exists a unique weighted least squares linear predictor and its parameters are given by*

$$\left(A^\top W A^\top \right)^{-1} A^\top W y.$$