



Why

What is the best linear predictor if we choose according to a particular norm.

Definition

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

$$g(y - Ax).$$

Weight matrix

Let $\|\cdot\|_W$ be the weighted norm for some postivie semidefinite weight matrix W . We want to find x to minimize

$$\|y - AX\|_W.$$

This problem is referred to by many authors as *weighted least squares* or the *weighted least squares problem*.

Diagonal weight matrix

A special case of norm weighted least squares with a weighted norm is the usual weighted least squares problem (see). Consider weighted least squares with weights $w \in \mathbf{R}^n$, $w \geq 0$. 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 and

$$\|y - Ax\|_W = \sum_{i=1}^n w_i (y_i - x^\top a_i)^2.$$

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

