



Why

Here's a simple idea. If the set of precepts is a vector space, use a predictor that is a linear transformation.¹

Definition

A *linear predictor* is a predictor which is linear in the precepts.

Example

Suppose that the set of precepts is \mathbf{R}^d , for some $d \in \mathbf{N}$.

Squared Loss

We have a dataset of pairs $(x^1, y^1), \dots, (x^n, y^n)$, and we want to select the predictor f to minimize. We want to find f linear to minimize $\sum_i \ell(f(x^i), y^i)$. For every linear function f there exists vector θ . We want to find f linear to minimize $\sum_i \ell(\theta^\top, y^i)$.

¹Future editions will expand on this why.

