Why

If we treat the parameters of a linear function as a random variable, an inductor for the predictor is equivalent to an estimator for the parameters.[1]

Definition

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Let $x : \Omega \to \mathbf{R}^d$. Define $g : \Omega \to (\mathbf{R}^d \to \mathbf{R})$ by $g(\omega)(a) = a^\top x(\omega)$, for $a \in \mathbf{R}^d$. In other words, for each outcome $\omega \in \Omega$, $g_\omega : \mathbf{R}^d \to \mathbf{R}$ is a linear function with parameters $x(\omega)$. g_ω is the function of interest.

Let $a^1, \dots, a^n \in \mathbf{R}^d$ a dataset with data matrix $A \in \mathbf{R}^{n \times d}$. Let $e : \Omega \to \mathbf{R}^n$ independent of x, and define $y : \Omega \to \mathbf{R}^n$ by

$$y = Ax + e.$$

In other words, $y_i = x^\top a^i + e_i$.

We call (x, A, e) a *probabilistic linear model*. Other terms include *linear model*, *statistical linear model*, *linear regression model*, *bayesian linear regression*, and *bayesian analysis of the linear model*.[2] We call x the parameters, A a *design*, e the *error* or *noise* vector, and y the *observation* vector.

One may want an estimator for the parameters x in terms of y or one may be modeling the function g and want to predict $g(a)$ for $a \in A$ not in the dataset.

---

[1]Future editions will offer further discussion.

[2]The word bayesian is in reference to treating the object of interest—x—as a random variable.

Inconsistency

In this model, the dataset is assumed to be inconsistent as a result of the random errors. In these cases, the error vector e may model a variety of sources of error ranging from inaccuracies in the measurements (or measurement devices) to systematic errors from the "inapproriateness" of the use of a linear predictor.[3] In this case the linear part is sometimes called the *deterministic effect* of the response on the input $a \in A$.

Moment assumptions

One route to be more specific about the underlying distribution of the random vector is give its mean and variance. It is common to give the mean of $\mathbf{E}(w)$

Mean and variance

Proposition 1. $\mathbf{E}(y) = A\mathbf{E}(x) + \mathbf{E}(w)$[4]

Proposition 2. $\mathrm{cov}((x, y)) = A\,\mathrm{cov}(x)A^\top + \mathrm{cov}\,e$[5]

A simple consequence is that, if x and

---

[3]Future editions will clarify and may excise this sentence.
[4]By linearity. Full account in future editions.
[5]Full account in future editions.