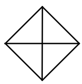

Why

There is a natural predictor corresponding to a normal linear model.

Definition

Let $(x : \Omega \to \mathbf{R}^d, A \in \mathbf{R}^{n \times d}, e : \Omega \to \mathbf{R}^n)$ be a normal linear model over the probability space $(\Omega, \mathcal{A}, \mathbf{P})$.

Predictive density

We are modeling $h_\omega : \mathbf{R}^d \to \mathbf{R}$ by $h_w(a) = x(\omega)^\top a$. The *predictive density* for a dataset $c^1, \dots, c^m \in \mathbf{R}^d$ is the conditional density of the random vector $(h_{(\cdot)}(c^1), \dots, h_{(\cdot)}(c^m))$ given y.

Proposition 1. *The predictive density for $c^1, \dots, c^m \in \mathbf{R}^d$ (with data matrix $C \in \mathbf{R}^{m \times d}$) is normal with mean*

$$g(a) = (C\Sigma_x A^\top)(A\Sigma_x A^\top + \Sigma_e)^{-1}\gamma.$$

and covariance

$$C\Sigma_x C^\top - C\Sigma_x A^\top (A\Sigma_x A^\top + \Sigma_e)^{-1} A\Sigma_x C^\top.$$

Proof. Define (as usual) $y : \Omega \to \mathbf{R}^n$ and $z : \Omega \to \mathbf{R}^m$ by

$$\begin{aligned} y &= Ax + e \\ z &= Cx. \end{aligned}$$

Recognize (x, y, z) as jointly normal, and use Normal Conditionals). □

Predictor

The *normal linear model predictor* or *normal linear model regressor* for the normal linear model (x, A, e) is the predictor which assigns to a new point $a \in \mathbf{R}^d$ the mean of the predictive density at a. That is, the predictor $g : \mathbf{R}^d \to \mathbf{R}$ defined by

$$g(a) = a^\top \Sigma_x A^\top (A\Sigma_x A^\top + \Sigma_e)^{-1}\gamma.$$

In the above we have substituted $a^\top$ for C. In the case of normal random vectors this corresponds with the MAP estimate and the MMSE estimate.[1] Notice that g is *linear* in its argument, a.

The use of a normal linear model predictor is often called *Bayesian linear regression*. The word Bayesian is used in reference to treating the parameters of the function, x, as a random variable.

[1]Future editions will have discussed this and include a reference to the sheet.