



## Why

We consider the probabilistic linear model in which all random variables are normal.

## Definition

A *normal linear model* is a probabilistic linear model in which the parameter and noise vectors have normal (Gaussian) densities. The model is also called the *Gaussian linear model* or the *linear model with Gaussian noise*.

Let  $(x : \Omega \rightarrow \mathbf{R}^d, A \in \mathbf{R}^{n \times d}, e : \Omega \rightarrow \mathbf{R}^n)$  be a probabilistic linear model over the probability space  $(\Omega, \mathcal{A}, \mathbf{P})$  in which  $x$  and  $e$  have normal densities. Recall that a probabilistic linear model has observation vector  $y : \Omega \rightarrow \mathbf{R}^n$  defined by

$$y = Ax + e.$$

## Conditional density of $x$ on $y$

Since  $x$  and  $e$  are normal and independent,  $y$  is normal.<sup>1</sup> Moreover, the random vector  $(x, y)$  is normal with covariance

$$\begin{pmatrix} \Sigma_x & \Sigma_x A^\top \\ A \Sigma_x & A \Sigma_x A^\top + \Sigma_e \end{pmatrix}.$$

So the conditional density (see **Normal Conditionals**) of  $g_{x|y}(\cdot, \gamma)$  is normal with mean

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

---

<sup>1</sup>Future editions will include an account.

and covariance

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

This density is sometimes called the posterior for the parameters given the observations. So the parameter posterior of the normal linear model is normal.

We can write the conditional mean as

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

and the conditional covariance as

$$(\Sigma_x^{-1} + A^\top \Sigma_e^{-1} A)^{-1} .^2$$

Very frequently we use these forms when  $d < n$ . In other words, in the case that we have fewer unknowns than measurements. In that case  $\Sigma_x$  is smaller than  $A \Sigma_x A^\top$ .

### Maximum conditional estimate of $x$

**Proposition 1.** *The maximum conditional estimate of  $x : \Omega \rightarrow \mathbf{R}^d$  given observed value  $\gamma \in \mathbf{R}^n$  of  $y : \Omega \rightarrow \mathbf{R}^n$  is the conditional mean*

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

---

<sup>2</sup>A proof will appear in future editions. Use the matrix inversion lemma or facts about inverses.

