



## Why

We want to estimate the weights of a linear function.<sup>1</sup>

## Definition

The *probabilistic linear model*; *linear model*; *linear regression*

Let  $(\Omega, \mathcal{A}, \mathbf{P})$  be a probability space. We have  $n$  precepts in  $\mathbf{R}^d$ . So let  $a^1, \dots, a^n \in \mathbf{R}^d$  with data matrix  $A \in \mathbf{R}^{n \times d}$ . We are modeling a relation between  $\mathbf{R}^d$  and  $\mathbf{R}$ .

Let  $x : \Omega \rightarrow \mathbf{R}^d$  and  $e : \Omega \rightarrow \mathbf{R}^n$  be independent random vectors with zero mean and covariances given by  $\Sigma_x$  and  $\Sigma_e$ , respectively. For each  $\omega \in \Omega$ , define the map  $f : \Omega \rightarrow (\mathbf{R}^d \rightarrow \mathbf{R})$  by  $f(\omega)(a) = \sum_j a_j^i x_j(\omega) + e_i(\omega)$ .

We call  $x$  the *signal*. We call  $e$  the *noise*. This class of models assumes the signal and noise are independent.

Define  $y : \Omega \rightarrow \mathbf{R}^n$  by  $y(\omega) = Ax(\omega) + e(\omega)$ . So,

$$y = Ax + e.$$

This is also called *bayesian linear regression* or the *bayesian analysis of the linear model*, in reference to the distribution on  $x$ .

## Mean and variance

**Proposition 1.**  $\mathbf{E}(y) = A\mathbf{E}(x) + \mathbf{E}(w)^2$

---

<sup>1</sup>Future editions will include this.

<sup>2</sup>By linearity. Full account in future editions.

**Proposition 2.**  $\mathbf{cov}((x, y)) = A \mathbf{cov}(x) A^\top + \mathbf{cov} e^3$

---

<sup>3</sup>Full account in future editions.



