



## Why

It is natural to embed a dataset.

## Definition

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  $(\Omega, \mathcal{A}, \mathbf{P})$ . Let  $\phi : \mathbf{R}^d \rightarrow \mathbf{R}^{d'}$  be a feature map.

We call the sequence  $(x, A, e, \phi)$  a *featurized probabilistic linear model* (also *embedded probabilistic linear model*). We interpret the model as a random field  $h : \Omega \rightarrow (\mathbf{R}^d \rightarrow \mathbf{R})$  which is a linear function of the features

$$h_\omega(a) = \phi(a)^\top x(\omega).$$

Denote the data matrix of the embedded feature vectors by  $\phi(A)$ . In other words,  $\phi(A) \in \mathbf{R}^{n \times d'}$  is a matrix whose rows are feature vectors. Then  $(x, A, e, \phi)$  corresponds to the probabilistic linear model  $(x, \phi(A), e)$ .

## Normal case

In the normal (Gaussian) case, the parameter posterior  $g_{x|y}(\cdot, \gamma)$  is a normal density with mean

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

and covariance

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

The predictive density for  $a \in \mathbf{R}^d$  is normal with mean

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

and covariance

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

So the *featurized linear regressor* is the predictor  $h : \mathbf{R}^d \rightarrow \mathbf{R}$  defined by

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



