Regression

Ha Phuong

September 29, 2021

1 Ex1

We assume that the target variable t is given by a deterministic function y(x, w) with additive Gaussian noise so that

$$t = y(x,w) + \in with \in \approx N(0, \beta^{-1})$$

$$=> p(t|x,w,\beta) = N((t|y(x,w), \beta^{-1}))$$

We have

$$y(x,w) = w^T \dot{\phi}(x_n)$$

with $\phi(xn)$ is basic function

$$=> p(t|x,w,\beta) = N((t|w^T \dot{\phi}(x_n), \beta^{-1}))$$

We will drop the explicit x from expressions (because this is supervised learning problems)

$$=> p(t|w,\beta) = N((t|w^T \dot{\phi}(x_n), \beta^{-1}))$$

Taking the logarithm of the likelihood function

$$=> lnp(t|w,) = \sum_{n=1}^{N} lnN(t_n|w^T \phi(xn), \beta^{-1})$$

$$=> lnp(t|w,) = \frac{N}{2} ln\beta - \frac{N}{2} ln(2\pi) - \beta \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2$$

Maximization of the likelihood function under a conditional Gaussian noise distribution for a linear model is equivalent to minimizing a sum-of-squares error function

$$=> \nabla lnp(t|w,) = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n)) \phi(x_n)^T$$

$$=> 0 = \sum_{n=1}^{N} t_n \phi(x_n)^T - w^T (\sum_{n=1}^{N} \phi(x_n) \phi(x_n)^) T$$

$$=> w_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t$$

Where $\Phi = \begin{pmatrix} \phi_0(x1) & \phi_1(x1) & \cdots & \phi_{M-1}(x1) \\ \phi_0(x2) & \phi_1(x2) & \cdots & \phi_{M-1}(x2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(x_N) & \phi_1(x_N) & \cdots & \phi_{M-1}(x_N) \end{pmatrix}$