Bayesian Regression

Mon Apr 12

Linear Discriminant Analysis

Regression

Discriminative model $p(y|x)$

vs.

Generative model

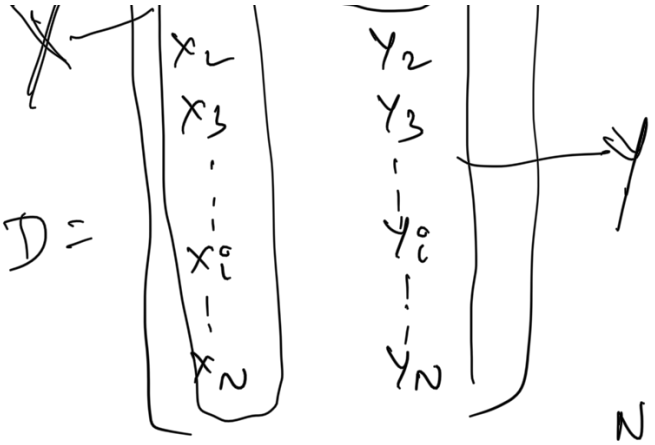
$p(y)$ $p(x|y)$

$w - ?$

$y|x,w = N(w^T x, \sigma^2)$

σ^2 → scalar

$[(x_1) \quad (y_1)]$



$$L(D|w) = \prod_{i=1}^{N} p(y_i | x_i, w)$$

$$LL(D|w) = \sum_{i=1}^{N} \log p(y_i | x_i, w)$$

$$= \sum_{i=1}^{N} \log \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (y_i - w^T x_i)^2 \right] \right]$$

$$= \sum_{i=1}^{N} \left[-\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} (y_i - w^T x_i)^2 \right]$$

$$= -\frac{N}{2}\log 2\pi - N\log\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - w^T x_i$$

Find (w, σ^2) that maximizes $LL(D|w)$

$$\frac{\partial LL(D|w)}{\partial} = 0 \quad \Big| \quad \frac{\partial LL(D|w)}{} = 0$$

∂w | $\partial \sigma$

$$\hat{w}_{MLE} = (X^T X)^{-1} X^T y$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{N} (y - X\hat{w}_{MLE})^T (y - X\hat{w}_{MLE})$$

Putting a prior on w

w is a $(D+1)$ length vector

$$p(w) \sim \mathcal{N}(w \mid \mu_0, \Sigma_0)$$

$$p(w \mid D) = \frac{p(D \mid w)\, p(w)}{\int_{w'} p(D \mid w')\, p(w')\, dw'}$$

Posterior ~~of~~ of w will also be a Gaussian.

$$p(w) \sim \mathcal{N}(w \mid 0, \tau^2 I)$$

τ^2 → scalar

~~of~~

Special but often-used prior on w

Posterior:

$$\bar{w} = \left(X^T X + \frac{\sigma^2}{\tau^2} I\right)^{-1} X^T y$$

$$\Sigma = \sigma^2 \left(X^T X + \frac{\sigma^2}{\tau^2} I\right)^{-1}$$

Think ridge regression