Note Title 4/27/18 1/22/2000

$$\{(\bar{x}_1, y_1) \ldots (\bar{x}_p, y_p)\} \quad x \in R^N$$

$$P_r(y_p / \bar{x}_p, \bar{\theta}) = \mathrm{Norm}_{y_p}[b + \bar{x}_p^T \bar{w}, \sigma^2]$$

$$\tilde{w} = (\tilde{X}\tilde{X}^T)^{-1} \tilde{X} \bar{y}$$

$$\sigma^2 = \frac{(\tilde{y} - \tilde{X}^T \tilde{w})^2}{P}$$

for new data $\bar{x}^*$

$$\bar{x}^{*T} \tilde{w} = \tilde{x}^{*T} (\tilde{X}\tilde{X}^T)^{-1} \tilde{X} \bar{y}$$

Bayesian Linear Regression

The rules of probability

sum rule : $p(X) = \sum_Y \underline{p(X, Y)}$ [marginalization]

product rule : $p(X, Y) = p(Y/X)\, p(X)$

$$\left. \begin{array}{l} p(X,Y) = p(Y/X)\,p(X) \\ p(X, Y) = p(Y, X) = p(X/Y)\, p(Y) \end{array} \right\}$$

$$\Rightarrow p(Y/X)\, p(X) = p(X/Y)\, p(Y)$$

$$\Rightarrow p(Y/X) = \frac{p(X/Y)\, p(Y)}{p(X)}$$ Bayes theorem

Given $\{(\bar{x}_p, y_p)\}_{p=1}^{P}$

Likelihood

$$Pr(\bar{y}/\bar{X}, \tilde{w}) = Norm_{\bar{y}}[\tilde{X}^T\tilde{w}, \sigma^2 I]$$

a distribution on $\tilde{w}$

$$Pr(\tilde{w}) = Norm_{\tilde{w}}[0, \sigma_p^2 I]$$

Using Bayes' theorem

posterior : $$\underline{Pr(\tilde{w}/\tilde{X}, \bar{y})} = \frac{Pr(\bar{y}/\bar{X}, \tilde{w}) \cdot Pr(\tilde{w})}{Pr(\bar{y}/\tilde{X})}$$

Marginal & Conditional of Gaussians

$$p(x) = Norm_x(\mu, \Lambda^{-1})$$

$$p(y/x) = Norm_y(Ax+b, L^{-1})$$

Then (see Bishop)

$$p(y) = Norm_y(A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$$

$$\rightarrow \quad p(x/y) = Norm_x(\Sigma(A^T L(y-b) + \Lambda\mu), \Sigma)$$

$$\Sigma = (\Lambda + A^T L A)^{-1}$$

$$A^T = \tilde{X}$$
$$L = \sigma^{-2} I$$
$$\Lambda = \sigma_p^{-2} I$$
$$\mu = 0$$
$$b = 0$$

$$\boxed{\begin{gathered} Pr(\tilde{w} / \tilde{X}, \bar{y}) = Norm_{\tilde{w}}\left[\frac{1}{\sigma^2} B^{-1} \tilde{X} \bar{y},\ B^{-1}\right] \\ B = \frac{1}{\sigma^2} \tilde{X} \tilde{X}^T + \frac{1}{\sigma_p^2} I \end{gathered}}$$

How to draw inference

$$\rightarrow \bar{x}^* \rightarrow Pr(y^* / \bar{x}^*, \tilde{X}, \bar{y})$$
$$= \int Pr(y^* / \bar{x}^*, \tilde{w})\, Pr(\tilde{w} / \tilde{X}, \bar{y})\, d\tilde{w}$$
$$\vdots$$

$$\boxed{\begin{gathered} = Norm_{y^*}\left[\frac{1}{\sigma^2} \bar{x}^{*T} B^{-1} \tilde{X} \bar{y},\ \bar{x}^{*T} B^{-1} \bar{x}^* + \sigma^2\right] \\ B = \frac{1}{\sigma^2} \tilde{X} \tilde{X}^T + \frac{1}{\sigma_p^2} I \end{gathered}}$$

If we use the mean of the gaussian

Bayesian: $\frac{1}{\sigma^2} \bar{x}^{*T} B^{-1} \tilde{X} \bar{y} = \frac{1}{\sigma^2} \bar{x}^{*T} \left(\frac{1}{\sigma^2} \tilde{X} \tilde{X}^T + \frac{1}{\sigma_p^2} I\right) \tilde{X} \bar{y}$

ML: $\bar{x}^{*T} (\tilde{X} \tilde{X}^T)^{-1} \tilde{X} \bar{y}$

Deterministic Approach

Regularizer

$$\min_{\tilde{w}} \left\{ \frac{1}{\sigma^2} \| \bar{y} - \tilde{X}^T \tilde{w} \|_2^2 + \frac{1}{\sigma_p^2} \| \tilde{w} \|^2 \right\}$$

OR $$\min_{\tilde{w}} \left\{ \| \bar{y} - \tilde{X}^T \tilde{w} \|_2^2 + \frac{\sigma^2}{\sigma_p^2} \| \tilde{w} \|^2 \right\}$$

$$\nabla_{\tilde{w}} (\cdot) = \frac{1}{\sigma^2} \left(2 \tilde{X} \tilde{X}^T \tilde{w} - 2 \tilde{X} \bar{y} \right) + \frac{1}{\sigma_p^2} 2 \tilde{w} = 0$$

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

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

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