$Data = \{(x_i, y_i)\}_{i=1}^{N}$

$x_i \in \mathbb{R}^p, \; y_i \in \mathbb{R},$

$$X = (x_1 \; x_2 \cdots x_N)^T = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix}_{N \times p}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N \times 1}$$

Model:

$$\begin{cases} f(x) = w^T x = x^T w \\ y = f(x) + \varepsilon \end{cases}$$

x, y, ε are r.v

$\varepsilon \sim N(0, \sigma^2)$

Bayesian Method:

$$\begin{cases} \text{Inference: posterior}(w) \\ \text{Prediction: } x^* \to y^* \end{cases}$$

Bayesian Linear Regression

$$\text{Inference: } P(w|Data) = P(w|X,Y) = \frac{P(w, Y|X)}{P(Y|X)} = \frac{\overset{\text{likelihood}}{P(Y|w,x)} \cdot \overset{\text{prior}}{P(w)}}{\int P(Y|w,x) \cdot P(w)\, dw}$$

$$P(Y|w,x) = \prod_{i=1}^{N} P(y_i|w,x_i) = \prod_{i=1}^{N} N(y_i | w^T x_i, \sigma^2)$$

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

$$P(w) = N(0, \Sigma_p)$$

共轭：Gaussian 分布是自共轭的

$$\underbrace{P(w|Data)}_{\text{Gaussian}} \propto \underbrace{P(Y|w,X)}_{\text{Gaussian}} \cdot \underbrace{P(w)}_{\text{Gaussian}}$$

$$\propto \prod_{i=1}^{N} N(y_i | w^T x_i, \sigma^2) \cdot N(0, \Sigma_p)$$

$$N(\mu_w, \Sigma_w)$$

$$\begin{cases} \mu_w = ? \\ \Sigma_w = ? \end{cases}$$

Bayesian Linear Regression

Data: $\{(x_i, y_i)\}_{i=1}^{N}$, $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$.

Model: $\begin{cases} f(x) = w^T x = x^T w \\ y = f(x) + \varepsilon \end{cases}$

$\varepsilon \sim N(0, \sigma^2)$

Bayesian Method:

参数: w 不是未知的常量

w 是一个概率分布

① Inference: $P(w|Data) \rightarrow$ posterior

$$\underbrace{\underset{\downarrow}{P(w|Data)}}_{N(\mu_w, \Sigma_w)} \propto \underset{\downarrow}{\text{likelihood}} \times \underset{\uparrow}{\text{prior}}$$

$N(\mu_w, \Sigma_w)$ $\quad N(\Delta, \Delta)$ $\quad N(\Delta, \Delta)$

$\mu_w = ?$, $\Sigma_w = ?$

② Prediction: Given x^*, y^*?

$$P(y^*|Data, x^*) = \int_w \underbrace{P(y^*|w, Data, x^*)}_{P(y^*|w, x^*)} \cdot \underbrace{P(w|Data, x^*)}_{\underbrace{P(w|Data)}_{\text{posterior}}} dw$$

Bayesian Linear Regression

Inference: $\boxed{P(w|Data)}$

$P(w|Data) \propto P(Y|X,w) \cdot P(w)$

$P(Y|X,w) = \prod_{i=1}^{N} N(w^T x_i, \sigma^2)$

$P(w) = N(0, \Sigma_p)$

令 $P(w|Data) = N(\mu_w, \Sigma_w)$

求 μ_w, Σ_w.

$X = (x_1\ x_2 \cdots x_N)^T_{N \times p}$

Likelihood:

$$P(Y|X,w) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - w^T x_i)^2\right\}$$

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

$$= \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp\left\{-\frac{1}{2}(Y - Xw)^T \sigma^{-2} I (Y - Xw)\right\}$$

$$= N(Xw, \sigma^{-2} I)$$

$$P(w|Data) \propto N(Xw, \sigma^{2} I) \cdot N(0, \Sigma_p)$$

$$\propto \exp\left\{-\frac{1}{2}(Y - Xw)^T \sigma^{-2} I (Y - Xw)\right\} \cdot \exp\left\{-\frac{1}{2} w^T \Sigma_p^{-1} w\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}(Y^T - w^T X^T)(Y - Xw) - \frac{1}{2} w^T \Sigma_p^{-1} w\right\}$$

$$\to N(\mu_w, \Sigma_w) \qquad \mu_w = ? \quad \Sigma_w = ?$$

Bayesian Linear Regression

$$P(w|Data) \propto \exp\{-\frac{1}{2\sigma^2}(Y^T - w^TX^T)(Y - Xw) - \frac{1}{2} w^T \Sigma_p^{-1} w\}$$

$$= \exp\{-\frac{1}{2\sigma^2}(Y^TY - 2Y^TXw \cancel{- w^TX^TY} + w^TX^TXw) - \frac{1}{2} w^T \Sigma_p^{-1} w\}$$

二次项：

$$-\frac{1}{2\sigma^2} w^TX^TXw - \frac{1}{2} w^T \Sigma_p^{-1} w = -\frac{1}{2}\left(w^T \underbrace{(\sigma^{-2} X^TX + \Sigma_p^{-1})}_{\Sigma_w^{-1} = A} w\right)$$

一次项：

$$+\frac{1}{\cancel{2}\sigma^2} \cdot (+\cancel{2})\, Y^TXw = \underbrace{\sigma^{-2} Y^TX}\, w$$

$$\mu_w^T A = \mu_w^T \Sigma_w^{-1} = \sigma^{-2} Y^TX$$

$$A\mu_w = \sigma^{-2} X^TY$$

$$\mu_w = \sigma^{-2} A^{-1} X^TY$$

$\therefore$ $P(w|Data) = N(\mu_w, \Sigma_w)$

$\mu_w = \sigma^{-2} A^{-1} X^TY$

$\Sigma_w = A^{-1}$ $\quad (A = \sigma^{-2} X^TX + \Sigma_p^{-1})$

Inference: $P(w|Data)$

$w|Data \sim N(\mu_w, \Sigma_w)$

$\mu_w = \sigma^{-2} A^{-1} X^T Y$

$\Sigma_w = A^{-1}$

$A = \sigma^{-2} X^T X + \Sigma_p^{-1}$

Prediction:

Given x^*, y^*.

Model: $f(x) = w^T x = x^T w$

$y = f(x) + \varepsilon$

$\varepsilon \sim N(0, \sigma^2)$

Bayesian Linear Regression $\longrightarrow$ Prediction

① $f(x^*)$, $P(f(x^*) | Data, x^*) = N(x^{*T}\mu_w, x^{*T}\Sigma_w x^*)$

noise-free

$f(x) = x^T w$

$f(x^*) = x^{*T} w$

$P(w) = N(\mu_w, \Sigma_w)$

$w \sim N(\mu_w, \Sigma_w)$

$x^{*T} w \sim N(x^{*T}\mu_w, x^{*T}\Sigma_w x^*)$

noise

② y^*, $y^* = f(x^*) + \varepsilon$

$\varepsilon \sim N(0, \sigma^2)$

$\therefore P(y^* | Data, x^*) = N(x^{*T}\mu_w, x^{*T}\Sigma_w x^* + \sigma^2)$