~~Till know~~

Frequentist Methods

Unknown θ
constant

$$\ell(\theta) = \log P(\text{data}; \theta)$$

$$-\ell(\theta) = \text{loss func}^n$$

$$\hat{\theta} = \arg\min_{\theta} -\ell(\theta)$$

$$= \arg\max_{\theta} \ell(\theta)$$

Now,

Bayesian Method

$\theta \rightarrow$ Random variable unobserved

$\theta \sim$ Prior distribution

X (Data) $\sim P(x|\theta)$

$$\underbrace{P(\theta|X)}_{\text{Posterior Distb}^n} = \frac{P(x|\theta) \cdot P(\theta)}{P(x)}$$

$$= \frac{P(x|\theta) \cdot P(\theta)}{\int P(x|\theta) \cdot P(\theta) \cdot d\theta}$$

In supervised ML

$\theta \perp X$

$\theta \sim$ Prior

$Y \sim P(Y|X, \theta)$

Posterior: $P(\theta|X, Y) = \dfrac{P(Y|X,\theta) \cdot P(\theta)}{P(Y|X)}$

Posterior Predictive Distn.

$$p(y_*|x, y, x_*) = \int p(y_*|x_*, \theta) . p(\theta|x, y) . d\theta$$

$$= E[p(y_*|x_*, \theta)]$$

$$\theta \sim p(\theta|x, y)$$

Parametric

$$y|x;\theta = \frac{1}{1+e^{-\theta^T x}}$$

BAYESIAN LINEAR REGRESSION

$$S = \{x^{(i)}, y^{(i)}\}_{i=1}^{n}$$

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

$$\epsilon^{(i)} \sim N(0, \sigma^2)$$

$$\theta \sim N(0, \tau^2 I)$$

$$\theta \in \mathbb{R}^d$$

$$x^{(i)} \in \mathbb{R}^d$$

$$p(\theta|S) \sim N\left(\frac{1}{\sigma^2} A^{-1} X^T \vec{y}, A^{-1}\right)$$

$$\text{where } A = \frac{1}{\sigma^2} X^T X + \frac{1}{\tau^2} I$$

Posterior Predictive

$$y_* | x_*, S \sim N\left(\frac{1}{\sigma^2} x_*^T A^{-1} X^T \vec{y},\ x_*^T A^{-1} x_* + \sigma^2\right)$$

Gaussian Processes

Prop. of M.V. Gaussians

① Normalization : $\int_x p(x; \mu, \Sigma) . dx = 1$

② Marginalization

$$x = \begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim N\left(\underbrace{\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}}_{\mu}, \underbrace{\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}}_{\Sigma}\right)$$

$$p(x_A) = \int_{x_B} p(x; \mu, \Sigma) . dx_B = N(\mu_A, \Sigma_{AA})$$

$$p(x_B) = \int_{x_A} p(x; \mu, \Sigma) . dx_A$$

③ Conditioning

$$x_A | x_B \sim N\left(\mu_A + \Sigma_{AB} \Sigma_{BB}^{-1} (x_B - \mu_B),\ \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}\right)$$

$$\begin{bmatrix} a \\ b \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}\right)$$

$$a \mid b \sim N\left(\mu_a + \frac{\rho\,\sigma_a \sigma_b}{\sigma_b^2}(b-\mu_b),\ \sigma_a^2 - \frac{\rho\,\sigma_a\sigma_b}{\sigma_b^2}\,\rho\,\sigma_a\sigma_b\right)$$

$$\sim N\left(\mu_a + \rho\cdot\sigma_a\left(\frac{b-\mu_b}{\sigma_b}\right),\ \sigma_a^2(1-\rho^2)\right)$$

④ <u>Summation</u>

$$x \sim N(\mu_1, \Sigma_1)$$
$$y \sim N(\mu_2, \Sigma_2)$$
$$x+y \sim N(\mu_1+\mu_2,\ \Sigma_1+\Sigma_2)$$

<u>M.V Gaussian</u> <u>Gaussian Process</u>

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}, \begin{bmatrix} \ \end{bmatrix}\right) \Leftrightarrow \begin{bmatrix} f \end{bmatrix} \sim GP\left(\begin{bmatrix} m \end{bmatrix}, \begin{bmatrix} K \end{bmatrix}\right)$$

Marginalize out all irrelevant examples
(Eg. not in training & not in test)

$$\begin{bmatrix} f(x^{(1)}) \\ \vdots \\ f(x^{(n)}) \\ f(x_*^{(1)}) \\ \vdots \\ f(x_*^{(n^*)}) \end{bmatrix} = N\left(\begin{bmatrix} 0 \end{bmatrix}, \begin{bmatrix} K(x^{(1)}, x^{(1)}), \dots K(x, x_*^{(1)(n)}) \\ \vdots \quad \vdots \end{bmatrix} \right)$$

↳ PSD
(Mercer Thm)

$$\begin{bmatrix} \vec{f} \\ \vec{f}_* \end{bmatrix} = N\left(0, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X & K(X_*, X_*) \end{bmatrix} \right)$$

$$y = f(x) + \varepsilon$$

$$\begin{bmatrix} y \\ y_* \end{bmatrix} = \begin{bmatrix} f \\ f_* \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \varepsilon_* \end{bmatrix}$$

$$\sim N\left(\begin{bmatrix} 0 \end{bmatrix}, \begin{bmatrix} K(X,X)+\sigma^2 I & K(X,X_*) \\ K(X,X_*) & K(X_*,X_*)+\sigma^2 I \end{bmatrix}\right)$$

$$\vec{\varepsilon} \sim N(0, \sigma^2 I)$$

Posterior Predictive

$$Y_* \mid Y, X, X_* \sim N(\mu_*, \Sigma_*)$$

$$\mu_* = K(X_*, X)\left[K(X,X)+\sigma^2 I\right]^{-1} Y$$

$$\Sigma_* = K(X_*, X_*) + \sigma^2 I$$

$$- K(X_*, X) \times \left[K(X,X)+\sigma^2 I\right]^{-1} K(X, X_*)$$