# (十九) 贝叶斯线性回归

$f(x) = w^T x$, $y = f(x) + \varepsilon = w^T x + \varepsilon$

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

LSE (最小二乘) → Regularized LSE (Lasso, Ridge)

LSE ↑ MLE (noise is Gaussian)

Regularized LSE ↓ MAP (noise is Gaussian)

$$w_{MLE} = \arg\max_w P(Data|w)$$

$$P(w|Data) \propto P(Data|w) P(w)$$

$$w_{MAP} = \arg\max_w P(w|Data)$$

$$= \arg\max_w P(Data|w) P(w)$$

$P(w)$ → $N(\mu_0, \Sigma_0)$

Inference:

$$P(w|Data) = P(w|X, Y) = \frac{P(w, Y|X)}{P(Y|X)} = \frac{P(w, Y|X)}{P(Y|w, X) P(w|X)} / P(Y|X)$$

$$= \frac{P(Y|w, X) P(w)}{\int P(Y|w, X) \cdot P(w) \, dw}$$

$P(w) = P(w|X)$ 因为 w 定好了分布.

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

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

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

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

Gaussian

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

共轭: Gaussian 分布自共轭