模型和超参的选取

① 平方指数类核函数

$k(x,x') = \sigma_f^2 \exp(-\frac{1}{2}(x-x')^T M (x-x')) + \sigma_n^2 \delta_{pq}$

$M_1 = l^{-2} I$　各向同性

当 $|x-x'| >> l$ ，$k(x,x') \sim 0$

$M_2 = \mathrm{diag}(l)^{-2}$　各向异性

当 $|x_i - x_i'| >> l_i$　$k(x,x') \sim 0$

$M_3 = \Lambda\Lambda^T + \mathrm{diag}(l)^{-2}$

$-\frac{1}{2}(x-x')^T\Lambda\Lambda^T(x-x') - \frac{1}{2}\sum_{i=1}\left(\frac{x_i - x_i'}{l_i}\right)^2$

$\Lambda = [u_1\ u_2 \cdots u_k]$

$-\frac{1}{2}\sum_i [(x-x')^T u_i]^2$　在 u_i 方向上的投影

如果 $(x-x')^T u_i >> 1$　$k(x,x') \sim 0$

② 贝叶斯模型选择

不同的模型 $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3 \cdots$

超参的选取 θ

模型回归参数 w

例子：高斯过程超参的选取

$$p(\theta \mid y, X) \propto p(y \mid \theta X)\, p_0(\theta)$$

$$p(y \mid \theta X) = \frac{1}{\sqrt{|2\pi K_y|}} e^{-\frac{1}{2} y^T K_y^{-1} y}$$

$$K_y = K(X, X) + \sigma_n^2 I$$

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

$$\log p(y \mid \theta X) = \underbrace{-\frac{1}{2} y^T K_y^{-1} y}_{\text{数据匹配}} \underbrace{- \frac{1}{2}\log|K_y|}_{\text{复杂度}} - \frac{n}{2}\log 2\pi$$

数据匹配误差小，复杂度小

$$K_y = \sigma_f^2 \left[e^{-\frac{1}{2l^2}(x_i - x_j)^2} \right] + \sigma^2 I$$

$l \uparrow \quad |K_y| \downarrow$ 复杂度变小

③ 交叉验证

留一交叉验证

$$p(y \mid X\, y_{-i}\, \theta) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y-\mu_i)^2/2\sigma_i^2}$$

$$\mu_i = k_{-i} K_{-i}^{-1} y_{-i}$$

$$\sigma_i = k(x_i, x_i) - k_{-i} K_{-i}^{-1} k_{-i}^T$$

$k_{-i} = [k_{i1} \quad k_{i2} \cdots \quad k_{in}]^T$ 不含 k_{ii}

$K_{-i} = [k_{j,k}] \quad j,k \neq i$

$$L = \sum_{i=1}^{n} \log P(y_i \mid X \; y_{-i} \; \theta)$$

假设我们有 K^{-1}

$$\begin{bmatrix} K_{-i} & k_{-i} \\ k_{-i}^T & k(x_i, x_i) \end{bmatrix} \qquad (i=n)$$

$$\begin{bmatrix} A & b \\ b^T & d \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1} b s^{-1} b^T A^{-1} & -A^{-1} b s^{-1} \\ -s^{-1} b^T A^{-1} & s^{-1} \end{bmatrix}$$

$$s = d - b^T A b \qquad (K^{-1} y)_n = -s^{-1} \underbrace{b^T A^{-1} y_{-n}}_{\mu_n} + s^{-1} y_n$$

$$\Rightarrow \sigma_i = 1/[K^{-1}]_{ii}$$

$$\mu_i = y_i - [K^{-1} y]_i / [K^{-1}]_{ii}$$

$$L = \sum_i -\frac{1}{2} \log \sigma_i^2 - \frac{(y_i - \mu_i)^2}{2\sigma_i^2} - \frac{1}{2} \log(2\pi)$$