第三次作业

冯海桐-522031910557

2024 年 11 月 17 日

1

$$
\begin{aligned}
p(\beta \mid y,x) &= \frac{p(y \mid x,\beta)p(\beta)}{p(y \mid x)} \\
&= \frac{1}{p(y \mid x)} N(x^T\beta, \sigma^2 I) N(0, \tau I) \\
&= \frac{1}{p(y \mid x)} \frac{1}{(2\pi)^n(\sigma^2\tau)^{n/2}} \exp(-\frac{1}{2}\frac{\|y-X\beta\|^2}{\sigma^2} - \frac{1}{2}\frac{\|\beta\|^2}{\tau})
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\hat{\beta} &= \arg\max_{\beta} p(\beta \mid y,x) \\
&= \arg\min_{\beta} (\frac{1}{2}\frac{\|y-X\beta\|^2}{\sigma^2} + \frac{1}{2}\frac{\|\beta\|^2}{\tau})
\end{aligned} \tag{2}
$$

与岭回归公式比较，可得

$$
\lambda = \frac{\sigma^2}{\tau} \tag{3}
$$

2

$$
\begin{aligned}
\langle K(.,x_i), f\rangle_{H_k} &= \sum_{t=1}^{\infty} \frac{1}{\gamma_t}\gamma_t\phi_t(x_i)\cdot a_t \\
&= \sum_{t=1}^{\infty} a_t\phi_t(x_i) \\
&= f(x_i)
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\langle K(.,x_i), K(.,x_j)\rangle_{H_k} &= \sum_{t=1}^{\infty} \frac{1}{\gamma_t}\gamma_t\phi_t(x_i)\cdot \gamma_t\phi_t(x_j) \\
&= \sum_{t=1}^{\infty} \gamma_t\phi_t(x_i)\phi_t(x_j) \\
&= K(x_i, x_j)
\end{aligned} \tag{5}
$$

3

3.1

设 $J(\beta)=\frac{1}{2}(y-X\beta)^TW(y-X\beta)$
其中 W 是一个对角矩阵，对角线元素为 $K_\lambda(x_0,x_i)$

$$\nabla_\beta J(\beta)=X^TW(y-X\beta)=0 \tag{6}$$

推出

$$\beta=(X^TWX)^{-1}X^TWy \tag{7}$$

3.2

$$\begin{aligned}\hat{f}(x_0)&=\begin{bmatrix}1\ x_0^T\end{bmatrix}\hat{\beta}(x_0)\\&=\begin{bmatrix}1\ x_0^T\end{bmatrix}(X^TWX)^{-1}X^TWy\\&=\sum_{i=1}^{N}l_i(x_0)y_i\\&=l(x_0)^Ty\end{aligned} \tag{8}$$

所以

$$l(x_0)^T=\begin{bmatrix}1\ x_0^T\end{bmatrix}(X^TWX)^{-1}X^TW \tag{9}$$

$$\begin{aligned}l(x_0)^TX&=\begin{bmatrix}1\ x_0^T\end{bmatrix}(X^TWX)^{-1}X^TWX\\&=\begin{bmatrix}1\ x_0^T\end{bmatrix}\end{aligned} \tag{10}$$

考虑到矩阵 X 的形式如下：

$$X=\begin{bmatrix}1 & x_1^T\\1 & x_2^T\\\vdots & \vdots\\1 & x_N^T\end{bmatrix} \tag{11}$$

带入可得到

$$l(x_0)^T\mathbf{1}=1 \tag{12}$$

$$[x_1\ x_2\ \cdots\ x_N]\,l(x_0)=x_0 \tag{13}$$

其中 $\mathbf{1}$ 是一个 N 维的全 1 向量。
展开后得到：

$$\sum_{i=1}^{N}l_i(x_0)=1 \tag{14}$$

$$\sum_{i=1}^{N}x_il_i(x_0)=x_0 \tag{15}$$

4

4.1

$$
\begin{aligned}
&\min_{w,b}\ \frac{1}{2}\left\|w\right\|^2 \\
&\text{s.t. } y_i(w^Tx_i+b)\geq 1
\end{aligned}
\tag{16}
$$

4.2

令

$$
L(w,b,\alpha)=\frac{1}{2}\left\|w\right\|^2-\sum_{i=1}^{N}\alpha_i(y_i(w^Tx_i+b)-1) \tag{17}
$$

$$
\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0 \tag{18}
$$

$$
\frac{\partial L}{\partial b}=-\sum_{i=1}^{N}\alpha_i y_i=0 \tag{19}
$$

代入得

$$
\begin{aligned}
&\max_{\alpha}\ \sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j \\
&\text{s.t. } \sum_{i=1}^{N}\alpha_i y_i=0 \\
&\alpha_i\geq 0
\end{aligned}
\tag{20}
$$

超平面是在数据空间中将数据分成两个部分的决策边界。而支持向量是位于最接近超平面的点。它们是定义超平面最重要的数据点，因为它们决定了决策边界的位置。几何上，支持向量到超平面的距离是最小的。

4.3

$$
\begin{aligned}
&\alpha_7=0 \\
&\alpha_8=0.7 \\
&\alpha_9=0
\end{aligned}
\tag{21}
$$

5

$$
\begin{aligned}
Err(x_0) &= E\left[(Y - \hat{f}(x_0))^2 \mid X = x_0\right] \\
&= E\left[(f(x_0) + \varepsilon - \hat{f}(x_0))^2\right] \\
&= E\left[(f(x_0) - \hat{f}(x_0))^2\right] + E\left[\varepsilon^2\right] \\
&= E\left[f(x_0)^2 - 2f(x_0)\hat{f}(x_0) + \hat{f}(x_0)^2\right] + \sigma^2 \\
&= E\left[f(x_0)^2\right] - 2E\left[f(x_0)\hat{f}(x_0)\right] + E\left[\hat{f}(x_0)^2\right] + \sigma^2 \\
&= f(x_0)^2 - 2f(x_0)E\hat{f}(x_0) + E\left[\hat{f}(x_0)^2\right] + \sigma^2 \\
&= \left[E\hat{f}(x_0)\right]^2 - 2f(x_0)E\hat{f}(x_0) + f(x_0)^2 + E\left[\hat{f}(x_0)^2\right] - 2E\left[\hat{f}(x_0)E\hat{f}(x_0)\right] + \left[E\hat{f}(x_0)\right]^2 + \sigma^2 \\
&= \left[E\hat{f}(x_0) - f(x_0)\right]^2 + E\left[\hat{f}(x_0) - E\hat{f}(x_0)\right]^2 + \sigma^2 \\
&= Bias^2 + Var + \sigma^2
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
\omega &= E_y\left[Err_{in} - \overline{err}\right] \\
&= E_y\left[\frac{1}{N}\sum_{i=1}^{N} E_{Y^{new}}\left[Y_i^{new} - \hat{y}_i\right]^2 - \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[Y_i^{new} - \hat{y}_i\right]^2 - (y_i - \hat{y}_i)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[(Y_i^{new})^2 - 2Y_i^{new}\hat{y}_i + \hat{y}_i^2\right] - y_i^2 + 2y_i\hat{y}_i - \hat{y}_i^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[(Y_i^{new})^2\right] - 2\hat{y}_i E_{Y^{new}} Y_i^{new} + \hat{y}_i^2 - y_i^2 + 2y_i\hat{y}_i - \hat{y}_i^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_y(y_i^2) - 2\hat{y}_i E_y y_i - y_i^2 + 2y_i\hat{y}_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[E_y(y_i^2) - 2E_y\hat{y}_i E_y y_i - E_y(y_i^2) + 2E_y(y_i\hat{y}_i)\right] \\
&= \frac{2}{N}\sum_{i=1}^{N}\left[E_y(y_i\hat{y}_i) - E_y\hat{y}_i E_y y_i\right] \\
&= \frac{2}{N}\sum_{i=1}^{N} cov(y_i, \hat{y}_i)
\end{aligned}
\tag{23}
$$