第四次作业

冯海桐-522031910557

2024 年 11 月 26 日

1

1.1 T

1.2 F

1.3 F

1.4 F

2

训练次数，卷积核大小、个数步长，pooling 大小，全连接层大小与个数，激活函数，损失函数，优化器。

随机 dropout。

3

3.1

$$p(y \mid x, \beta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{(y - f(x, \beta))^2}{2\sigma_\varepsilon^2}\right) \tag{1}$$

$$\begin{aligned} p(Y \mid X, \beta) &= \prod_{i=1}^{N} p(y_i \mid x_i, \beta) \\ &= \frac{1}{(2\pi)^{N/2}\sigma_\varepsilon^N} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2\right) \end{aligned} \tag{2}$$

因此

$$\begin{aligned} arg \max_\beta p(Y \mid X, \beta) &= arg \max_\beta \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2\right) \\ &= arg \min_\beta \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2 \end{aligned} \tag{3}$$

上述的 N 表示训练样本的数量。

3.2

$$\begin{aligned}
p(\beta \mid y, x) &= \frac{p(y \mid x, \beta)p(\beta)}{p(y \mid x)} \\
&\propto p(y \mid x, \beta)p(\beta) \\
&= \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{(y - f(x, \beta))^2}{2\sigma_\varepsilon^2}\right) \frac{1}{(2\pi)^{N/2}\tau^N} \exp\left(-\frac{\|\beta\|^2}{2\tau^2}\right)
\end{aligned} \tag{4}$$

上述的 N 表示 β 的维度。在没有显式的给出 $f(x, \beta)$ 的形式时，无法直接计算 β 的后验分布，也就无从得知是否为高斯分布。

$$\begin{aligned}
p(\beta \mid y, x) &\propto \exp\left(-\frac{(y - f(x, \beta))^2}{2\sigma_\varepsilon^2} - \frac{\|\beta\|^2}{2\tau^2}\right) \\
&= \exp\left(-\frac{1}{2\sigma_\varepsilon^2}(y - \sum_{j=1}^{N} \beta_j h_j(x))^2 - \frac{\sum_{i=1}^{N} \beta_i^2}{2\tau^2}\right) \\
&= \exp\left(-\frac{1}{2}(\frac{1}{\sigma_\varepsilon^2}(y^2 + \sum_{j=1}^{N}(\beta_j^2 h_j^2(x) - 2y\beta_j h_j(x))) + \frac{1}{\tau^2}\sum_{i=1}^{N}\beta_i^2)\right) \\
&= \exp\left(-\frac{1}{2}(\frac{1}{\sigma_\varepsilon^2}y^2 + \sum_{j=1}^{N}\frac{(\tau^2 h_j^2(x) + \sigma_\varepsilon^2)}{\sigma_\varepsilon^2\tau^2}(\beta_j^2 - 2\frac{\sigma_\varepsilon^2\tau^2 y h_j(x)}{\tau^2 h_j^2(x) + \sigma_\varepsilon^2}\beta_j))\right)
\end{aligned} \tag{5}$$

由此可见

$$\beta_j \sim (\sigma_\varepsilon^2\tau^2 y h_j(x)/(\tau^2 h_j^2(x) + \sigma_\varepsilon^2), \sigma_\varepsilon\tau/\sqrt{\tau^2 h_j^2(x) + \sigma_\varepsilon^2}) \tag{6}$$

4

$$
\begin{aligned}
op &= E_y\left[Err_{in} - \overline{err}\right] \\
&= E_y\left[\frac{1}{N}\sum_{i=1}^{N} E_{Y^{new}}\left[Y_i^{new} - \hat{y}_i\right]^2 - \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[Y_i^{new} - \hat{y}_i\right]^2 - (y_i - \hat{y}_i)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[(Y_i^{new})^2 - 2Y_i^{new}\hat{y}_i + \hat{y}_i^2\right] - y_i^2 + 2y_i\hat{y}_i - \hat{y}_i^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_{Y^{new}}\left[(Y_i^{new})^2\right] - 2\hat{y}_i E_{Y^{new}} Y_i^{new} + \hat{y}_i^2 - y_i^2 + 2y_i\hat{y}_i - \hat{y}_i^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_y\left[E_y(y_i^2) - 2\hat{y}_i E_y y_i - y_i^2 + 2y_i\hat{y}_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[E_y(y_i^2) - 2E_y\hat{y}_i E_y y_i - E_y(y_i^2) + 2E_y(y_i\hat{y}_i)\right] \\
&= \frac{2}{N}\sum_{i=1}^{N}\left[E_y(y_i\hat{y}_i) - E_y\hat{y}_i E_y y_i\right] \\
&= \frac{2}{N}\sum_{i=1}^{N} cov(y_i, \hat{y}_i)
\end{aligned} \tag{7}
$$

在线性回归中

$$
\begin{aligned}
op &= \frac{2}{N}\sum_{i=1}^{N} cov(y_i, \hat{y}_i) \\
&= \frac{2}{N}\sum_{i=1}^{N}\left[E_y(y_i\hat{y}_i) - E_y\hat{y}_i E_y y_i\right] \\
&= \frac{2}{N}\sum_{i=1}^{N}\left[E_y(y_i\overline{x_i}^T(X^TX)^{-1}X^TY) - E_y(\overline{x_i}^T(X^TX)^{-1}X^TY)E_y y_i\right] \\
&= \frac{2}{N}\sum_{i=1}^{N}\overline{x_i}^T(X^TX)^{-1}X^T\left[E_y(y_iY) - E_yYE_y y_i\right]
\end{aligned} \tag{8}
$$

由于 $Y = [y_1\ y_2\ \cdots\ y_d]^T$, 当 $i \neq j$ 时，y_i 与 y_j 独立，因此 $E_y(y_iy_j) = E_y y_i E_y y_j$。
所以有

$$
\begin{aligned}
E_y(y_iY) - E_yYE_y y_i &= \left[0\ 0\ 0\ \cdots\ E_y y_i^2 - (E_y y_i)^2\ \cdots\ 0\right]^T \\
&= \left[0\ 0\ 0\ \cdots\ \sigma_\varepsilon^2\ \cdots\ 0\right]^T
\end{aligned} \tag{9}
$$

因此

$$
\begin{aligned}
op &= \frac{2}{N}\sum_{i=1}^{N}\overline{x_i}^T(X^TX)^{-1}X^T\left[E_y(y_iY)-E_yYE_yy_i\right] \\
&= \frac{2}{N}\sum_{i=1}^{N}\overline{x_i}^T(X^TX)^{-1}X^T\begin{bmatrix}0\;0\;0\;\cdots\;\sigma_\varepsilon^2\;\cdots\;0\end{bmatrix}^T \\
&= \frac{2\sigma_\varepsilon^2}{N}\sum_{i=1}^{N}\overline{x_i}^T(X^TX)^{-1}\overline{x_i} \\
&= \frac{2\sigma_\varepsilon^2}{N}tr\left(X(X^TX)^{-1}X^T\right) \\
&= \frac{2\sigma_\varepsilon^2}{N}tr\left(X^TX(X^TX)^{-1}\right) \\
&= \frac{2\sigma_\varepsilon^2}{N}tr\left(I\right) \\
&= \frac{2d}{N}\sigma_\varepsilon^2
\end{aligned} \tag{10}
$$