MATH455(Statistical Learning)- Solution2

魏上傑

2023-04-20

Quiz Problem: Note that this exercise shows that the OLS estimate is essentially the MLE.

sol:

We use the following notation to prove the result.

$X = \begin{bmatrix} 1 & x_{11} & ... & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & ... & x_{Np} \end{bmatrix}$ be a $N \times (p+1)$ matrix, Y be a $N \times 1$ vector, and β be a $(p+1) \times 1$ vector.

Notice that

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{n=1}^{N} L(y, f(x)) \quad = \arg\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{n=1}^{N} -log(p(x,y)) \; = \arg\max_{f \in \mathcal{F}} \frac{1}{N} \sum_{n=1}^{N} log(p(x,y))$$

That is, we want to maximize the log-likelihood function.

Also, note that

$$log(p(x,y)) \propto p(x,y) = p(y|x)p(x) \propto p(y|x)$$

$$Y|X \sim N(X^T\beta, \sigma^2)$$

$$\implies p(Y|X) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(Y - X\beta)^2}{2\sigma^2}) \propto -(Y - X\beta)^2$$

Hence, maximize $p(Y|X) \iff$ minimize $(Y - X\beta)^2$

$$\frac{\partial}{\partial \beta} = 0 : -2X^T(Y - X\beta) = 0$$

$$\implies X^TY - X^TX\beta = 0 \implies \hat{\beta} = (X^TX)^{-1}X^TY$$

Selected Problems:

7.

$$
\begin{aligned}
&Y \sim N(X\beta, \sigma^2 I) \\
&\implies (X^T X)^{-1} X^T Y \sim N((X^T X)^{-1} X^T X\beta, (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1}) \\
&\implies \hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})
\end{aligned}
$$

We use the fact that σ^2 is scalar.

8.

$$
\begin{aligned}
||y - X\beta||^2 &= (y - X\beta)^T (y - X\beta) \\
&= (y^T - \beta^T X^T)(y - X\beta) \\
&= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta \\
&= y^T y - 2y^T X\beta + \beta^T X^T X\beta
\end{aligned}
$$

Note that $y^T X\beta$ is a scalar, so $y^T X\beta = (y^T X\beta)^T$

9.

$$\frac{d}{d\beta}(y^T y) = 0$$

This is because $y^T y$ does not depend on β

10.

$$\frac{d}{d\beta}(y^T X\beta) = X^T y$$

Check the dimension: β is $(p+1) \times 1$, and $X^T y$ is $((p+1) \times N) \times (N \times 1) = (p+1) \times 1$

11.

$$\frac{d}{d\beta}(\beta^T X^T X\beta) = 2(X^T X)\beta$$

Check the dimension: β is (p+1)x1, and $(X^T X)\beta$ is $((p+1) \times N) \times (N \times (p+1)) \times ((p+1) \times 1) = (p+1) \times 1$