Quiz2

2022年11月1日 星期二 10:21

(1) $y = Ax+n$，x为未知参数，(A, y)为已知数据，
$n \in N(0, Q)$，求x的极大似然估计

解：① $p(y|A,x) \sim N(Ax, Q)$

$$\ln p(y|A,x) = \underbrace{\eta}_{与x无关的常数} + (y-Ax)^T Q^{-1} (y-Ax) \left(-\frac{1}{2}\right)$$

要使 $\ln p(y|A,x)$ 最大，需使 $(y-Ax)^T Q^{-1} (y-Ax)$ 最小，
这就引出最小二乘。

② $L = (y-Ax)^T Q^{-1} (y-Ax)$，记 $e = y-Ax$

$$\frac{\partial L}{\partial x} = \frac{\partial e^T Q^{-1} e}{\partial x} = \frac{\partial e^T Q^{-1} e}{\partial e} \frac{\partial e}{\partial x} = \underset{1\times q}{2e^T} \underset{q\times q}{Q^{-1}} \cdot \underset{q\times n}{(-A)}$$

③ 由必要条件 $\frac{\partial L}{\partial x}(x^*) = 0 \Rightarrow$

$$2(y-Ax)^T Q^{-1}(-A) = 0$$

$$A^T (Q^{-1})^T (y - Ax^*) = 0$$

$$A^T (Q^{-1})^T y = A^T (Q^{-1})^T A x^*$$

$$x^* = (A^T (Q^{-1})^T A)^{-1} A^T (Q^{-1})^T y$$

若A可逆，进一步化简有 $x^* = A^{-1} Q^T (A^T)^{-1} A^T (Q^T)^{-1} y$

$= A^{-1} y$ （与Q无关）

(2) A与x均为未知参数，求A与x的极大后验估计。

⓪ 记 $f(A)$ 为 A.reshape(−1, 1)，即flatten后的向量。

设 $f(A) \sim N(\mu_A, \Sigma_A)$ $\quad x \sim N(\mu_x, \Sigma_x)$

① $\ln p(A,x|y) = \ln p(y|A,x) + \ln p(A,x) + \eta$

$= \ln p(y|A,x) + \ln p(A) + \ln p(x) + \eta$ （独立）

$= \left(-\frac{1}{2}\right)(y-Ax)^T Q^{-1} (y-Ax) + \left(-\frac{1}{2}\right)(f(A)-\mu_A)^T$
$\Sigma_A^{-1} (f(A)-\mu_A) + \left(-\frac{1}{2}\right) (x-\mu_x)^T \Sigma_x^{-1} (x-\mu_x)$

使后验 $\ln p(A,x|y)$ 极大，需使 损失函数L最小，

$$L = (y-Ax)^T Q^{-1} (y-Ax) + (f(A)-\mu_A)^T \Sigma_A^{-1} (f(A)-\mu_A) + (x-\mu_x)^T \Sigma_x^{-1} (x-\mu_x)$$

② $\frac{\partial L}{\partial A}(A^* | 已知x) = 0 \quad \frac{\partial L}{\partial x}(x^* | 已知A) = 0$

(1) $\frac{\partial L}{\partial x} = 2(y-Ax)^T Q^{-1} (-A) + 2(x-\mu_x)^T \Sigma_x^{-1}$

$\Rightarrow x^* = (\Sigma_x^{-1} + A^T R^{-1} A)^{-1} A^T R^{-1} y$

(2) 由于 $y^T = x^T A^T + n^T$，因此有（对称性）

$\Rightarrow A^* = (y R^{-1} x^T)(\Sigma_A^{-1} + x R^{-1} x^T)^{-1}$

③ $A_0 \to x_1 \to A_1 \to \cdots \to x_n \to A_n$

当 x_n, A_n 满足一定条件后停止迭代

(3) ① Linear Regression $y = Ax+n$
如(1)所示。

② Linear Classification (Binary)

1° Linear Discriminant Analysis

project $\underset{q\times m}{A} \to \underset{q\times 1}{A'}$

数据量 $\quad$ 二分类，

such that 二类数据之间的方差最大。

随后在直线上找边界进行二分类。

$A' = Aw \quad w \propto (m_1 - m_2)$

其中 $\underset{m\times 1}{m_i}$ 为属于i类的数据的平均值。

s_1，s_2 为类内协方差。

构造 $J(w) = \frac{(m_2-m_1)^2}{s_1^2+s_2^2} = \frac{w^T S_B w}{w^T S_w w}$

得 $w \propto S_w^{-1}(m_2 - m_1)$

2° Logistic Regression

对 $\underbrace{[\ln y - \ln(1-y)]}_{\text{logits}} = Ax+n$

进行回归

但不能直接把 y处理为logits再回归，
这样会出现ln0，做不了。

只能够移到右边变为sigmoid

$$y = \sigma(Ax+n)$$