2025 期中

题目一

简述用AIC , C_p方法对线性回归模型的选择过程

C_p 公式:

$$C_p \triangleq \frac{SSE(p)}{s^2} - [\,n - 2(p+1)\,]$$

C_p 本质上用

$$C_p \approx p+1$$

来判断模型是否既不过度拟合也不过度简化。

$\text{AIC} \triangleq \max \text{loglikelihood}(k) - k$ (k 为模型中待估参数的数量)

- **概率密度函数（pdf）**：

$$f(Y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \|Y - X_p\beta_p\|_2^2 \right\}$$

- **对数似然函数**:

$$\text{loglikelihood} = -\frac{1}{2\sigma^2}\|Y - X_p\beta_p\|_2^2 - \frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2$$

- 回归系数的MLE：$\hat{\beta}_p = (X_p^T X_p)^{-1} X_p^T Y$
- 方差的MLE：$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\|Y - X_p\hat{\beta}_p\|_2^2 = \frac{SSE(p)}{n}$

将MLE代入对数似然，整理得:

$$\max \text{loglikelihood} = -\frac{n}{2} - \frac{n}{2}\log 2\pi - \frac{n}{2}\log SSE(p) + \frac{n}{2}\log n$$

根据AIC定义 $\text{AIC} \triangleq \max \text{loglikelihood} - k$（此处 $k = p+1$，即 β_p 的维度 + 方差参数），代入并化简（消去与变量子集无关的常数项后）：

$$\text{AIC} \triangleq -\frac{n}{2}\log SSE(p) - p$$

应用：寻找使AIC最大的变量子集 X_p，以此平衡模型拟合度与复杂度。

题目二

已知线性模型 $Y = X\beta + \varepsilon$，$\hat{\beta}$ 是 β 的极大似然估计（MLE），满足：

$$P\left(\max_c \frac{|c'(\hat{\beta}-\beta)|^2}{c'(X^TX)^{-1}c} \geq M_\alpha\right) = \alpha$$

（其中 c 为任意非零向量）（hint：利用柯西施瓦茨（Cauchy-Schwarz）定理）求 M_α。

对矩阵 $\boldsymbol{X}'\boldsymbol{X}$ 做谱分解（实对称矩阵可正交对角化）：

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$$

其中 $\boldsymbol{U}$ 是正交矩阵（$\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$），$\boldsymbol{D}$ 是对角矩阵（对角线为 $\boldsymbol{X}'\boldsymbol{X}$ 的特征值）。

由此，$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}'$，因此：

$$\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}'\boldsymbol{c}$$

令 $\boldsymbol{v} = \boldsymbol{D}^{-1/2}\boldsymbol{U}'\boldsymbol{c}$（即 $\boldsymbol{D}^{1/2}\boldsymbol{v} = \boldsymbol{U}'\boldsymbol{c}$），则上式可改写为：

$$\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c} = \boldsymbol{v}'\boldsymbol{v}$$

原分式的分子为 $|\boldsymbol{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|^2 = [\boldsymbol{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2$，代入 $\boldsymbol{c} = \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{v}$ 得：

$$[\boldsymbol{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2 = [\boldsymbol{v}'\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2$$

因此，原分式可表示为：

$$\frac{[\boldsymbol{v}'\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2}{s^2\boldsymbol{v}'\boldsymbol{v}}$$

根据柯西施瓦茨定理：对任意向量 $\boldsymbol{a}, \boldsymbol{b}$，有 $|\boldsymbol{a}'\boldsymbol{b}| \leq \|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|$。

令 $\boldsymbol{a}=\boldsymbol{v}$，$\boldsymbol{b}=\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$，则:

$$|\boldsymbol{v}'\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})| \leq \|\boldsymbol{v}\| \cdot \|\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|$$

两边平方后除以 $\boldsymbol{v}'\boldsymbol{v}=\|\boldsymbol{v}\|^2$，得:

$$\frac{[\boldsymbol{v}'\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2}{\boldsymbol{v}'\boldsymbol{v}} \leq \|\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|^2$$

因此，原分式满足:

$$\frac{[\boldsymbol{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2}{s^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}} \leq \frac{\|\boldsymbol{D}^{1/2}\boldsymbol{U}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|^2}{s^2}$$

已知 $\hat{\boldsymbol{\beta}}$ 是 MLE，其分布为:

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$

构造统计量 $\frac{1}{\sigma}(\boldsymbol{X}'\boldsymbol{X})^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$，其中 $(\boldsymbol{X}'\boldsymbol{X})^{1/2}=\boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}'$，易知:

$$\frac{1}{\sigma}(\boldsymbol{X}'\boldsymbol{X})^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$

因此，该统计量的平方范数服从卡方分布:

$$\left\|\frac{1}{\sigma}(\boldsymbol{X}'\boldsymbol{X})^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\|^2 = \frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2} \sim \chi^2_{k+1}$$

（注：$k+1$ 是 $\boldsymbol{\beta}$ 的维度，即自变量个数+截距项）

记残差方差 $s^2=\frac{\mathrm{SSE}}{n-(k+1)}$，其中 $\mathrm{SSE} \sim \sigma^2\chi^2_{n-(k+1)}$，因此 $\frac{(n-(k+1))s^2}{\sigma^2} \sim \chi^2_{n-(k+1)}$。

构造F统计量（卡方变量除以自由度后相除）:

$$\frac{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2(k+1)}}{\frac{(n-(k+1))s^2}{\sigma^2(n-(k+1))}} = \frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{s^2(k+1)} \sim F_{k+1,n-(k+1)}$$

结合步骤3的不等式，原分式的最大值满足:

$$\max_{\boldsymbol{c}} \frac{|\boldsymbol{c}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|^2}{s^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}} \le \frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{s^2}$$

而由F分布的分位数定义，$P\left(\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{s^2(k+1)} \ge F_{\alpha;k+1,n-(k+1)}\right) = \alpha$，即：

$$P\left(\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{s^2} \ge (k+1)F_{\alpha;k+1,n-(k+1)}\right) = \alpha$$

因此，最终得到：

$$M_\alpha = (k+1)F_{\alpha;k+1,n-(k+1)}$$

题目三

3、Model：$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$，$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$，记少了第$i$次观测后得到的$\hat{\boldsymbol{\beta}}_{(i)}$，已知：$\boldsymbol{H}$，$\boldsymbol{x}_i$，$R_i$，$R_i = y_i - \hat{y}_i$

1）请用$\boldsymbol{H}$，$\boldsymbol{x}_i$，X，R_i 给出$\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}$的表示形式

2）求$\frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)})^2$

已知条件

- 原线性模型：$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$，原参数估计为 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$；
- 删去第i次观测后，设计矩阵为$\boldsymbol{X}_{(i)}$（保留$n-1$次观测）、响应向量为$\boldsymbol{Y}_{(i)}$，对应参数估计为$\hat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{X}_{(i)}^T\boldsymbol{Y}_{(i)}$；
- 残差：$R_i = y_i - \hat{y}_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$；
- 帽子矩阵$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$，其对角线元素$H_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$。

$\hat{\beta}_{(i)} - \hat{\beta}$的表示形式

矩阵逆的差满足：$\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} = -\boldsymbol{A}^{-1}(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{B}^{-1}$。
令$\boldsymbol{A} = \boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}$、$\boldsymbol{B} = \boldsymbol{X}^T\boldsymbol{X}$，则：

$$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}-(\boldsymbol{X}^T\boldsymbol{X})^{-1}=-(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}-\boldsymbol{X}^T\boldsymbol{X}\right)(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

由于$\boldsymbol{X}$比$\boldsymbol{X}_{(i)}$多了第i行$\boldsymbol{x}_i^T$，因此:

$$\boldsymbol{X}^T\boldsymbol{X}=\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}+\boldsymbol{x}_i\boldsymbol{x}_i^T$$

移项得:

$$\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}-\boldsymbol{X}^T\boldsymbol{X}=-\boldsymbol{x}_i\boldsymbol{x}_i^T$$

将$\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}-\boldsymbol{X}^T\boldsymbol{X}=-\boldsymbol{x}_i\boldsymbol{x}_i^T$代入逆的差公式，得:

$$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}-(\boldsymbol{X}^T\boldsymbol{X})^{-1}=(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

$\boldsymbol{X}^T\boldsymbol{Y}$可拆分为“第$i$次观测的贡献”与“剩余$n-1$次观测的贡献”之和:

$$\boldsymbol{X}^T\boldsymbol{Y}=\boldsymbol{x}_iy_i+\boldsymbol{X}_{(i)}^T\boldsymbol{Y}_{(i)}$$

因此:

$$\boldsymbol{X}_{(i)}^T\boldsymbol{Y}_{(i)}=\boldsymbol{X}^T\boldsymbol{Y}-\boldsymbol{x}_iy_i$$

对$\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\right]$变形:
由$\boldsymbol{X}^T\boldsymbol{X}=\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}+\boldsymbol{x}_i\boldsymbol{x}_i^T$，两边左乘$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}$、右乘$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$，得:

$$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{x}_i\left(1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\right)=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$$

因此:

$$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{x}_i=\frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i}{1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i}$$

将$\boldsymbol{X}_{(i)}^T\boldsymbol{Y}_{(i)}=\boldsymbol{X}^T\boldsymbol{Y}-\boldsymbol{x}_iy_i$代入$\hat{\boldsymbol{\beta}}_{(i)}$的定义:

$$\hat{\boldsymbol{\beta}}_{(i)} = \left[(\boldsymbol{X}^T\boldsymbol{X})^{-1} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}}{1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i} \right] \left(\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{x}_i y_i \right)$$

结合$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$，展开$\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}$：

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}} &= \left[(\boldsymbol{X}^T\boldsymbol{X})^{-1} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}}{1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i} \right] \left(\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{x}_i y_i \right) - \hat{\boldsymbol{\beta}} \\
&= \hat{\boldsymbol{\beta}} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}{1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i} - \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i}{1-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i} - \hat{\boldsymbol{\beta}}
\end{aligned}$$

合并分子项（注意$R_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$，即$\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} - y_i = -R_i$），并代入$H_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$，最终化简得：

$$\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}} = \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i R_i}{1-H_{ii}}$$

计算$\frac{1}{n}\sum_{i=1}^n (y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)})^2$

将$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i R_i}{1-H_{ii}}$代入$y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)}$：

$$\begin{aligned}
y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)} &= y_i - \boldsymbol{x}_i^T \left(\hat{\boldsymbol{\beta}} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i R_i}{1-H_{ii}} \right) \\
&= (y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}) - \frac{\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i R_i}{1-H_{ii}} \\
&= R_i - \frac{H_{ii}R_i}{1-H_{ii}} \\
&= \frac{R_i(1-H_{ii}) - H_{ii}R_i}{1-H_{ii}} \\
&= \frac{R_i}{1-H_{ii}}
\end{aligned}$$

因此：

$$(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)})^2 = \frac{R_i^2}{(1-H_{ii})^2}$$

求和并取平均，最终结果为：

$$
\frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(i)})^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i^2}{(1-H_{ii})^2}
$$

题目四

推导logistic似然函数的log likelihood；写出Ridge形式下的对数似然函数（即给似然函数加了一个罚因子），并推导其对应的更新公式。

Logistic 似然函数的 log-likelihood

设 Logistic 回归中，第 i 个观测的响应变量 y_i（取值为 0 或 1）服从伯努利分布，其概率为：

$$
p_i = P(y_i = 1 \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{\exp\left\{\boldsymbol{x}_i^T\boldsymbol{\beta}\right\}}{1+\exp\left\{\boldsymbol{x}_i^T\boldsymbol{\beta}\right\}}
$$

$$
1 - p_i = P(y_i = 0 \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{1}{1+\exp\left\{\boldsymbol{x}_i^T\boldsymbol{\beta}\right\}}
$$

似然函数:
独立观测下，样本的似然函数为各观测概率的乘积:

$$
L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left[p_i^{y_i}\cdot(1-p_i)^{1-y_i}\right]
$$

对数似然函数:
对似然函数取自然对数（将乘积转化为求和，简化计算）：

$$
\begin{aligned}
\ell(\boldsymbol{\beta}) &= \log L(\boldsymbol{\beta}) \\
&= \sum_{i=1}^{n}\left[y_i\log p_i + (1-y_i)\log(1-p_i)\right]
\end{aligned}
$$

代入 p_i 化简:
将 $p_i = \frac{\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}}$ 代入，利用 $\log p_i = \boldsymbol{x}_i^T\boldsymbol{\beta} - \log\left(1+\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}\right)$、$\log(1-p_i) = -\log\left(1+\exp\{\boldsymbol{x}_i^T\boldsymbol{\beta}\}\right)$，化简得:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \left(\boldsymbol{x}_i^T \boldsymbol{\beta} - \log \left(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\} \right) \right) + (1 - y_i) \left(- \log \left(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\} \right) \right) \right]$$
$$= \sum_{i=1}^{n} \left[y_i \boldsymbol{x}_i^T \boldsymbol{\beta} - \log \left(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\} \right) \right]$$

Ridge 形式下的对数似然与更新公式

Ridge 正则化通过添加 L_2 惩罚项控制参数复杂度，适用于 Logistic 回归的过拟合问题。

Ridge 形式的对数似然函数:
在原对数似然基础上，添加 $\frac{\lambda}{2}\|\boldsymbol{\beta}\|^2$（$\lambda \geq 0$ 为正则化参数，$\|\boldsymbol{\beta}\|^2 = \boldsymbol{\beta}^T \boldsymbol{\beta}$ 为 L_2 范数平方）的惩罚项，得:

$$\ell(\lambda, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \boldsymbol{x}_i^T \boldsymbol{\beta} - \log \left(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\} \right) \right] - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2$$

一阶导数（梯度）：
对 $\ell(\lambda, \boldsymbol{\beta})$ 关于 $\boldsymbol{\beta}$ 求导（梯度）：

- 原对数似然部分的导数: $\nabla \left(\sum_{i=1}^n y_i \boldsymbol{x}_i^T \boldsymbol{\beta} \right) = \boldsymbol{X}^T \boldsymbol{y}$; $\nabla \left(- \sum_{i=1}^n \log(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\}) \right) = -\boldsymbol{X}^T \boldsymbol{p}$（其中 $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)^T$）。
- 惩罚项的导数: $\nabla \left(-\frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \right) = -\lambda \boldsymbol{\beta}$。

因此梯度为:

$$\nabla \ell(\lambda, \boldsymbol{\beta}) = \boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{X}^T \boldsymbol{p} - \lambda \boldsymbol{\beta}$$

二阶导数（Hessian 矩阵）：
对梯度继续关于 $\boldsymbol{\beta}$ 求导（Hessian 矩阵）：

- 原对数似然部分的二阶导数: $\nabla^2 \left(- \sum_{i=1}^n \log(1 + \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\}) \right) = -\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$（其中 $\boldsymbol{W} = \operatorname{diag}\{p_1(1-p_1), p_2(1-p_2), \ldots, p_n(1-p_n)\}$ 为对角矩阵）。
- 惩罚项的二阶导数: $\nabla^2 \left(-\frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \right) = -\lambda \boldsymbol{I}$（$\boldsymbol{I}$ 为单位矩阵）。

因此 Hessian 矩阵为:

$$\nabla^2 \ell(\lambda, \boldsymbol{\beta}) = -\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} - \lambda \boldsymbol{I}$$

更新公式（Newton-Raphson 迭代）：
Newton-Raphson 法通过“当前参数 - Hessian 逆 × 梯度”更新参数，即：

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}_0 - \left[\nabla^2 \ell(\lambda, \boldsymbol{\beta})\right]^{-1} \cdot \nabla \ell(\lambda, \boldsymbol{\beta})$$

将梯度和 Hessian 代入，注意 $-\left[\nabla^2 \ell(\lambda, \boldsymbol{\beta})\right]^{-1} = \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1}$，最终更新公式为：

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}_0 + \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1} \left(\boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{X}^T \boldsymbol{p}_0 - \lambda \boldsymbol{\beta}_0\right)$$

（注：$\boldsymbol{\beta}_0$ 为迭代初始值，$\boldsymbol{p}_0$、$\boldsymbol{W}$ 均基于 $\boldsymbol{\beta}_0$ 计算）

题目五

给定线性模型 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$，原假设 $H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$，似然比定义为：

$$\lambda = \frac{\max\limits_{H_0} L(\boldsymbol{\beta}_0, \boldsymbol{Y})}{\max\limits_{\Omega} L(\boldsymbol{\beta}, \boldsymbol{Y})} \quad \text{(likelihood ratio)}$$

其中 Ω 为 $\boldsymbol{\beta}$ 的无约束参数空间。求：

1）求 H_0 下的 $\max\limits_{H_0} L(\boldsymbol{\beta}_0, \boldsymbol{Y})$；

2）求似然比 λ 的表示形式；

3）求 $-2\log\lambda$ 的渐近分布（$n \to \infty$）。

1）求 H_0 下的 $\max\limits_{H_0} L(\boldsymbol{\beta}_0, \boldsymbol{Y})$

在 $H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$ 下，最大化似然函数等价于最小化残差平方和 $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$（似然函数的指数部分）。构造拉格朗日函数：

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{c})$$

其中 $\boldsymbol{\lambda}$ 为拉格朗日乘子。

对 $\boldsymbol{\beta}$ 求导并令其为0：

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0} \implies 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0} \tag{1}$$

约束条件：$\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$

记无约束下的OLS估计为 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$，将 $\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 代入式(1)：

$$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - \frac{1}{2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}$$

代入约束条件(2)：

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} - \frac{1}{2}\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{c}$$

解得拉格朗日乘子：

$$\boldsymbol{\lambda} = 2\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})$$

将 $\boldsymbol{\lambda}$ 代回 $\boldsymbol{\beta}$ 的表达式，得 H_0 下的约束估计：

$$\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})$$

在 H_0 下，σ^2 的极大似然估计为：

$$\hat{\sigma}_0^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)$$

将 $\hat{\boldsymbol{\beta}}_0$ 和 $\hat{\sigma}_0^2$ 代入似然函数，得 H_0 下的最大似然：

$$\max_{H_0} L(\boldsymbol{\beta}_0, \boldsymbol{Y}) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}}\exp\left\{-\frac{n}{2}\right\}$$

2）求似然比 λ 的表示形式

无约束下的最大似然（Ω 下）：

$$\max_{\Omega} L(\boldsymbol{\beta}, \boldsymbol{Y}) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp\left\{-\frac{n}{2}\right\}$$

其中 $\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ 为无约束下 σ^2 的极大似然估计。

似然比定义为“约束最大似然 / 无约束最大似然”，代入得：

$$\lambda = \frac{\max_{H_0} L}{\max_{\Omega} L} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$$

3）求 $-2\log\lambda$ 的渐近分布（$n \to \infty$）

由 $\lambda = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$，取对数得：

$$-2\log\lambda = n\log\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)$$

由约束估计的性质，约束残差平方和与无约束残差平方和满足：

$$(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})$$

因此：

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = 1 + \frac{(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})}{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}$$

当 $n \to \infty$ 时，$\frac{(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})}{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})} \to 0$，利用近似 $\log(1+x) \approx x$（$x \to 0$），得：

$$\log\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right) \approx \frac{(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T\left[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})}{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}$$

由无约束估计的渐近性质：$\hat{\boldsymbol{\beta}} \sim N_p\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)$，因此：

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c} \sim N_m\left(\boldsymbol{0}, \sigma^2\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right) \quad (\text{若 } H_0\text{成立})$$

进而：

$$(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T\left[\sigma^2\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T\right]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) \sim \chi^2(m)$$

同时，$\frac{1}{\sigma^2 n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \xrightarrow{p} 1$（依概率收敛），因此：

$$-2\log\lambda \xrightarrow{d} \chi^2(m) \quad (n \to \infty)$$

最终结论：当 $n \to \infty$ 时，$-2\log\lambda$ 渐近服从自由度为 m 的卡方分布 $\chi^2(m)$。