统计方法与机器学习

理论习题二

SCHOOL OF DATA SCIENCE & ENGINEERING
数据科学与工程学院

习题1

证明:

$$\hat{\sigma}^2 = \frac{1}{n-p-1} SS_E$$

是 σ^2 的无偏估计。

习题2

考虑一元线性回归模型

$$y = \beta_0 + \beta_1 x$$

现有数据 $\{(x_i, y_i): i = 1,2,\cdots,n\}$。对数据进行变换

$$\tilde{y}_i = \frac{y_i - c_1}{d_1},\ x_i = \frac{x_i - c_2}{d_2}, i = 1,2,\cdots,n$$

其中，c_1, c_2, d_1, d_2 为提前确定的常数。请完成以下任务：

(a) 试构建由原始数据和变换后数据得到的回归系数的最小二乘估计、总偏差平方和、回归平方和以及残差平方和之间的关系。

(b) 证明：由原始数据和变换后数据得到的 F 统计量的值保持不变。

习题3

3. 对给定的 n 组数据 $(x_i, y_i), i = 1, 2, \cdots, n$, 若我们关心的是 y 如何依赖 x 的取值而变动, 则可以建立回归方程

$$\hat{y} = a + bx$$

反之, 若我们关心的是 x 如何依赖 y 的取值而变动, 则可以建立另一个回归方程

$$\hat{x} = c + dy$$

试问这两条直线在直角坐标系中是否重合? 为什么? 若不重合, 它们有无交点? 若有, 试给出交点的坐标。

习题4

令

$$\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

是一个帽子矩阵，$\boldsymbol{I}$ 是单位矩阵。证明： $\boldsymbol{I}-\boldsymbol{H}$ 是对称幂等矩阵，并计算这个矩阵的秩。

习题5

5. 在一个多元线性回归模型中, 响应变量 y_i 的回归值为

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$$

$\boldsymbol{X}$ 是一个满秩矩阵, 证明: $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$。

在多元线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

中, 我们有数据 $\{(y_i, x_{i1}, x_{i2}, \cdots, x_{ip})\}_{i=1}^{n}$。我们可以得到最小二乘估计, 记为 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)^\top$。如果我们对 $y_1, y_2, \cdots, y_n$ 进行中心化, 对每一维自变量 $x_{1j}, x_{2j}, \cdots, x_{nj}$ 均进行了标准化, $j = 1, 2, \cdots, p$, 那么, 我们得到的最小二乘估计为 $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \cdots, \tilde{\beta}_p)^\top$。

请完成以下任务。

(a) 这两个估计 $\tilde{\beta}$ 和 $\hat{\beta}$ 之间有什么关系?

(b) 求 $\tilde{\beta}$ 的期望和方差。

习题6

6. 在单因子方差分析模型

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a, \quad j = 1, 2, \cdots, m$$

其中, ε_{ij} 是独立同分布的随机变量, 其分布为 $N(0, \sigma^2)$。我们观测到的数据为 $\{y_{ij}\}$。
请论证: 单因子方差分析模型可以看作一种多元线性回归模型。具体来说:

(a) 构造一个合适的设计矩阵 $\boldsymbol{X}$;

(b) 定义响应变量向量、回归参数向量、设计矩阵、误差向量, 并写出“数据版”的多元线性回归模型;

(c) 最小二乘法估计回归参数向量, 并与 μ_i 进行比较;

(d) 利用 F 检验, 对所构造对多元线性回归模型进行模型显著性检验, 并与方差分析的结果进行比较。