

作业要求

(1) 证明西瓜书上公式 (3.27) $\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left(-y_i \boldsymbol{\beta}^T \mathbf{x}_i + \log \left(1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i) \right) \right)$ 关于 $\boldsymbol{\beta}$ 的梯度 $\nabla \ell$ 和 Hessian 矩阵 \mathbf{H} 分别为

$$\begin{aligned}\nabla \ell &= \mathbf{X}^T (\boldsymbol{\mu} - \mathbf{y}), \\ \mathbf{H} &= \mathbf{X}^T \mathbf{S} \mathbf{X},\end{aligned}$$

其中

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix} \mathbf{x}_1, \dots, \mathbf{x}_n \end{bmatrix}^T, \\ \boldsymbol{\mu} &= \begin{bmatrix} \frac{1}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{x}_1)}, \dots, \frac{1}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{x}_n)} \end{bmatrix}^T, \\ \mathbf{y} &= [y_1, \dots, y_n]^T, \\ \mathbf{S} &= \text{diag}(\boldsymbol{\mu}).\end{aligned}$$

进一步证明矩阵 \mathbf{H} 是正定矩阵。

(2) 对于 Breast Cancer Wisconsin 数据集，根据上述公式，用 Python 实现梯度下降法和牛顿法估计 Logistic Regression 模型的参数 $\boldsymbol{\beta}$ (参考李航《统计学习方法》附录 A&B)。进一步对比计算得到的结果与作业 2 中由 sklearn 算出的结果之间的异同。