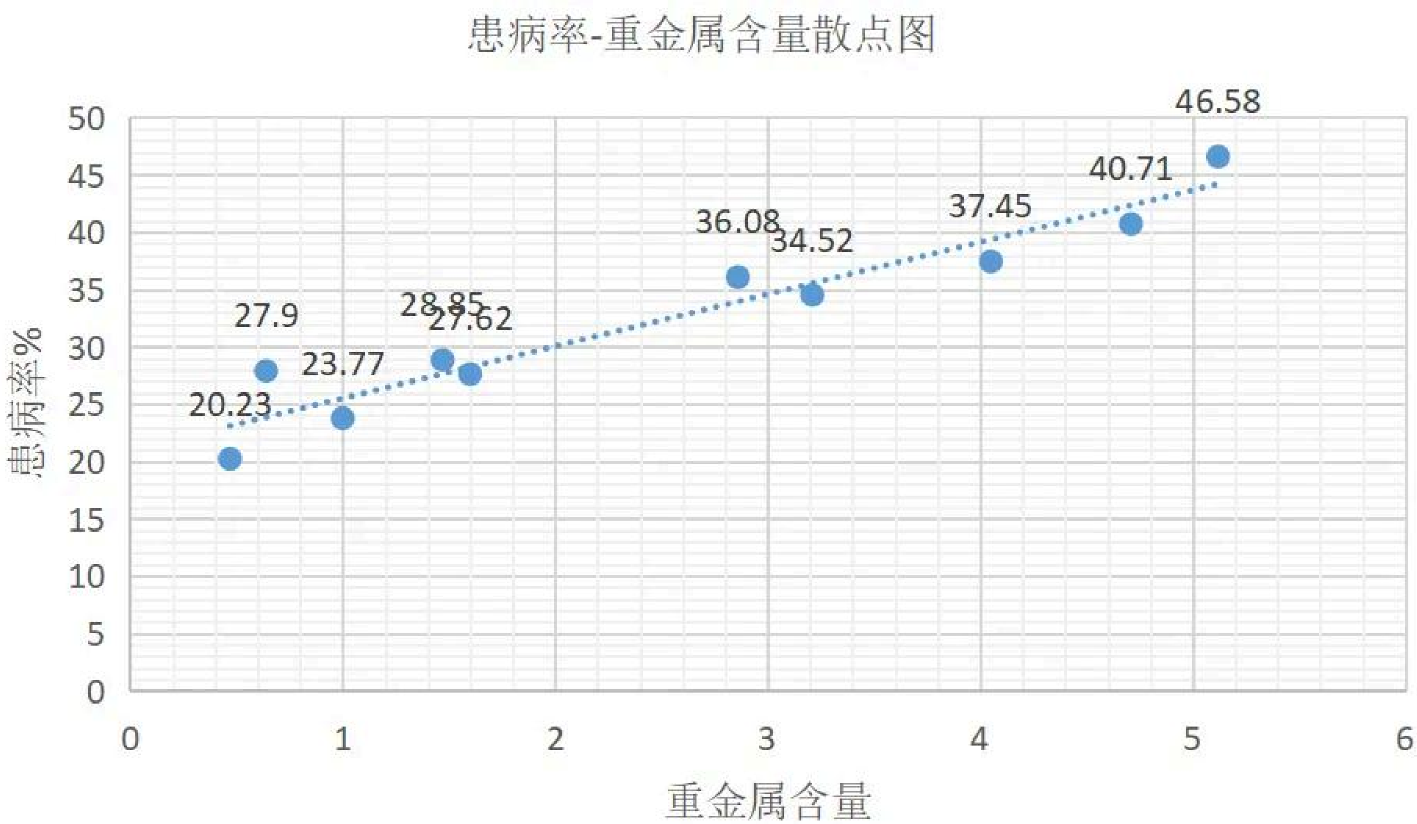
人工智能原理作业5

1

a)



b)

斜率: 4.54

截距: 20.97

r^2 的计算基于以下几个统计量：

- 总平方和（**Total Sum of Squares, SST**）：

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- 回归平方和（**Regression Sum of Squares, SSR**）：

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- 残差平方和（**Residual Sum of Squares, SSE**）：

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

通过以下公式计算：

$$r^2 = \frac{SSR}{SST} = \frac{545.14}{591.95} = 0.921$$

含义：

r^2的值范围从 0 到 1。值越接近 1，表示模型对数据的拟合越好，解释的变异越多，例如r^2 = 0.80可以解释为模型通过自变量解释了因变量 80% 的变异。

c)

$$MAE = \frac{1}{10}(|20.23 - 22.77| + |27.90 - 23.83| + ... + |46.58 - 44.33|)$$

$$MAE \approx 1.95$$

$$MSE = \frac{1}{10}((20.23 - 22.77)^2 + (27.90 - 23.83)^2 + ... + (46.58 - 44.33)^2)$$

$$MSE \approx 4.68$$

MAE和MSE的数值都比较小，表示模型对数据有着较好的拟合效果，误差水平控制得较好。

d)

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2$$

展开右边的平方项

$$SST = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

根据残差平方和最小这一估计条件，对回归方程系数进行求偏导：

$$\sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$$
$$\sum_{i=1}^{n} (y_i - \hat{y}_i) x_i = 0$$

$$\Sigma_{i=1}^{n}(y_i-\hat{y})(\hat{y}-\bar{y})=\Sigma_{i=1}^{n}(y_i-\hat{y})(wx_i+b)-\Sigma_{i=1}^{n}(y_i-\hat{y})\bar{y}=0$$

因此 SST=SSR+SSE

根据之前的计算：

SST=591.95

SSR=545.14

SSE=46.81

验证关系：

SST≈SSR+SSE，即591.95≈545.14+46.81

2

在二元逻辑回归的对数模型中，预测事件 y=1 的概率模型为

$$\log\left(\frac{P(y=1\mid x)}{1-P(y=1\mid x)}\right)=\beta^T x$$

从而

$$P(y=1\mid x)=\frac{e^{\beta^T x}}{1+e^{\beta^T x}}$$

我们可以定义第i个样本属于第k类的对数概率为：

$$\log P(y^{(i)}=k)=\beta_k^T x_i-\log Z$$

这里logZ是归一化因子，确保通过指数化和归一化后的所有类别概率之和为1：

$$Z=\sum_{j=1}^{K}e^{\beta_j^T x_i}$$

将对数形式转换为概率的形式，我们得到Softmax回归模型：

$$P(y^{(i)} = k) = \frac{e^{\beta_k^T x_i}}{Z}$$

3

a）

给定输入x1，Softmax模型首先计算每个类别的得分：

$$z_k = w_k x_1 + b_k$$

输出为类别k的概率为：

$$P(y = k \mid x_1) = \frac{e^{z_k}}{\sum_{j=1}^{l} e^{z_j}}$$

交叉熵损失：

$$L = -\sum_{k=1}^{l} y_{1k} \log P(y = k \mid x_1)$$

其中y1k表示当y1确实为类别k的时候取1，否则取0.

使用梯度下降求解该模型，要求损失L对zk的偏导数，我们先考虑 P(y = k|x1)的偏导数。根据链式法则:

$$\frac{\partial L}{\partial w_k} = \frac{\partial L}{\partial P(y=k|x_1)} \cdot \frac{\partial P(y=k|x_1)}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_k}$$

$$= x_1 \cdot (P(y = k \mid x_1) - y_{1k})$$

$$\frac{\partial L}{\partial b_k} = 1 \cdot (P(y = k \mid x_1) - y_{1k})$$

b)

计算出梯度后更新参数：

$$w_k = w_k - \alpha \cdot \frac{\partial L}{\partial w_k}$$

$$b_k = b_k - \alpha \cdot \frac{\partial L}{\partial b_k}$$