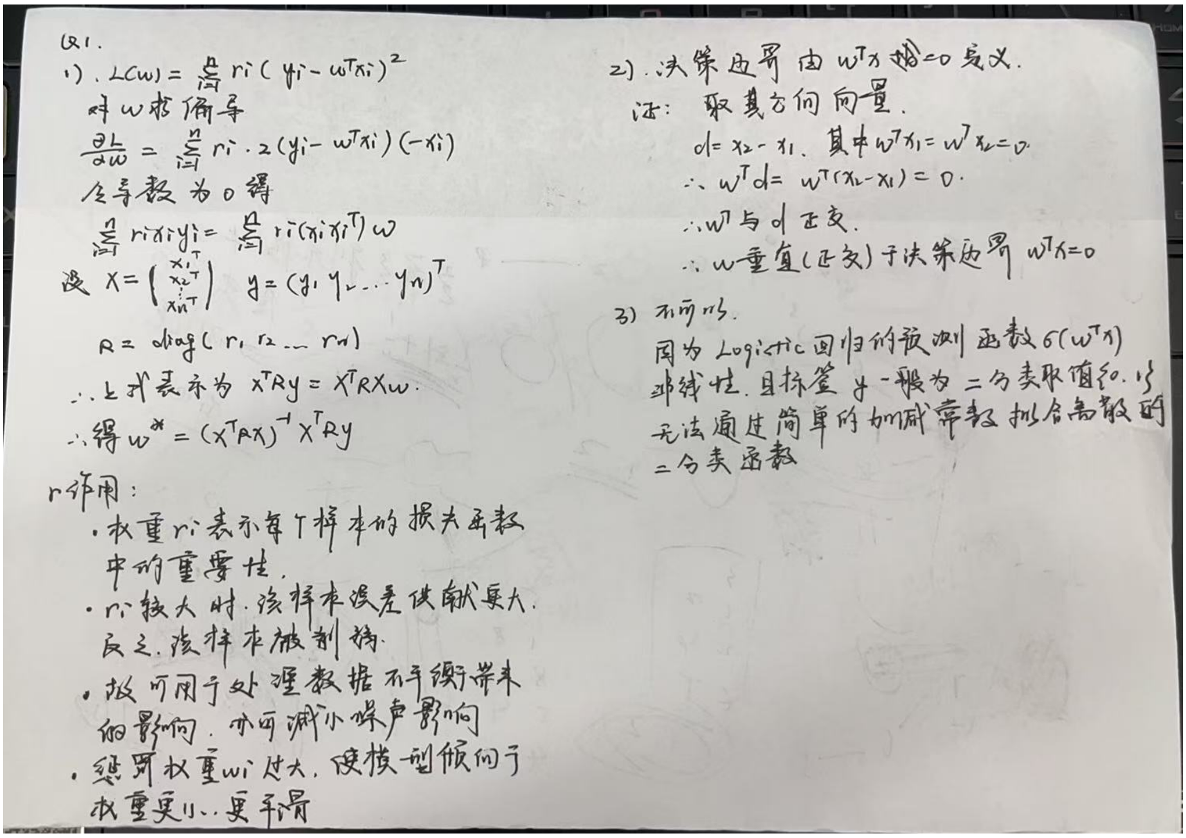
# A1-QA

## Q1



Q1.

1). $L(w)=\sum_{i=1}^{n} r_i(y_i-w^Tx_i)^2$

对 $w$ 求偏导

$\frac{\partial L}{\partial w}=\sum_{i=1}^{n} r_i\cdot 2(y_i-w^Tx_i)(-x_i)$

令导数为 0 得

$\sum_{i=1}^{n} r_i x_i y_i=\sum_{i=1}^{n} r_i(x_i x_i^T)w$

设 $X=\begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$ $y=(y_1, y_2 \ldots y_n)^T$

$R=\mathrm{diag}(r_1, r_2 \ldots r_n)$

∴ 上式表示为 $X^TRy=X^TRXw$.

∴ 得 $w^*=(X^TRX)^{-1}X^TRy$

r 作用：

- 权重 $r_i$ 表示每个样本的损失函数中的重要性.
- $r_i$ 较大时. 该样本误差贡献更大. 反之. 该样本被削弱.
- 故可用于处理数据不平衡带来的影响. 亦可减小噪声影响
- 惩罚权重 $w_i$ 过大, 使模型倾向于权重更小、更平滑

2). 决策边界由 $w^Tx$ ~~始~~ $=0$ 定义.

证： 取其方向向量.

$d=x_2-x_1$, 其中 $w^Tx_1=w^Tx_2=0$.

∴ $w^Td=w^T(x_2-x_1)=0$.

∴ $w^T$ 与 $d$ 正交.

∴ $w$ 垂直（正交）于决策边界 $w^Tx=0$

3) 不可以.

因为 Logistic 回归的预测函数 $\sigma(w^Tx)$ 非线性. 目标签 $y$ 一般为二分类取值{0,1} 无法通过简单的加减常数拟合离散的二分类函数

## Q2

1. 因为 $Y$ 在数据集固定的时候是一个常数，即 $\lambda T^TT$ 是一个常数

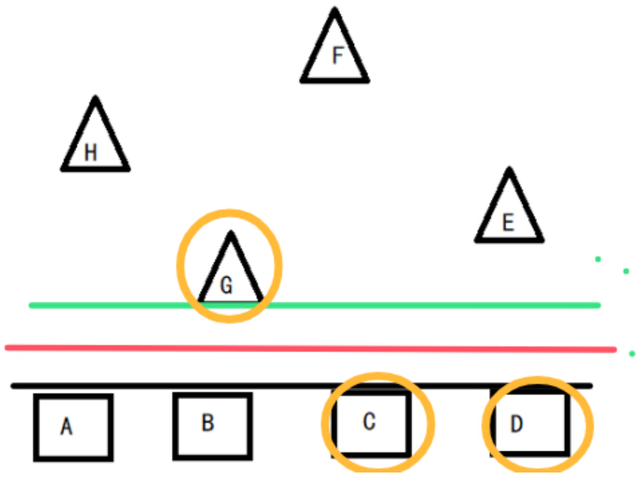
   其在优化过程中，无法通过梯度对参数优化起到作用，因此起不到正则化的作用

2. 对于 L2 正则化，其作用是在损失函数中增加 $\omega$ 的大小项，起到对 $\omega$ 过大时的惩罚，减小过拟合的风险

   若 $\lambda<0$，则效果恰好相反，$\omega$ 越大，反而可以得到更小的损失，无法达到限制 $\omega$ 的效果

## Q3

1.

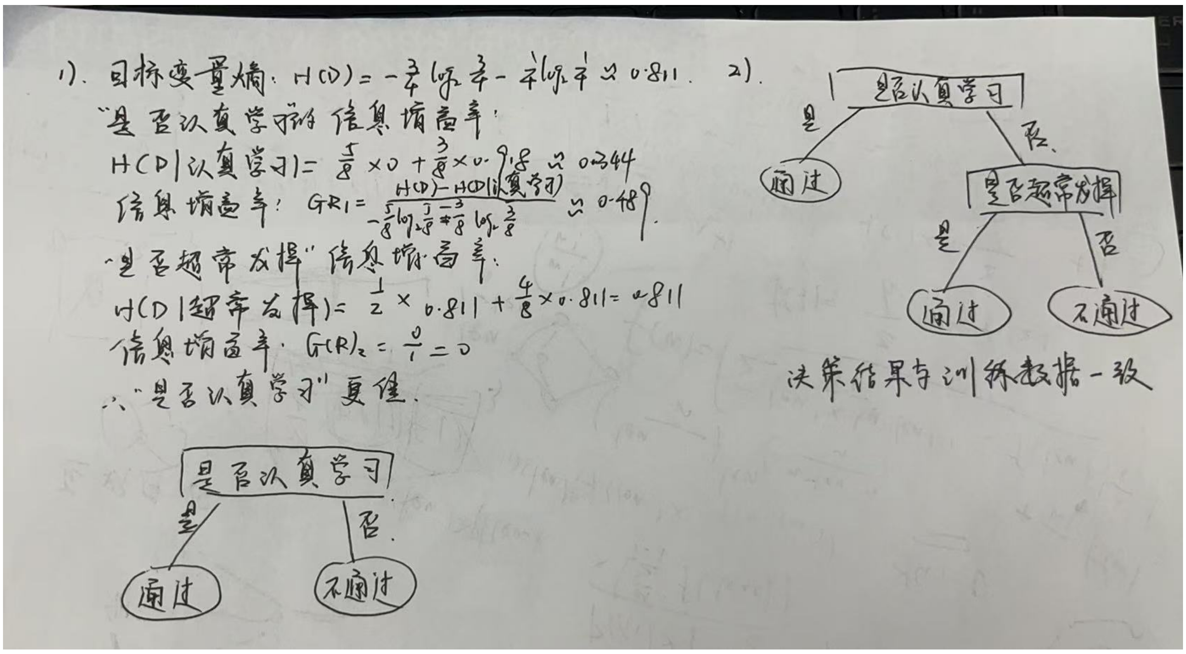



2. 在图中，所有的正方形都在直线下方，所有的三角形都在直线上方，因此训练错误率为 **0**

3. 根据 SVM 的理论，只有**支持向量**会映像最终的分类边界，对于其他非支持向量集合中的点，移除都不会改变间隔平面的解

   因此，只有移除这些支持向量才会导致分类边界发生变化，即 G 或者 C、D

# Q4



1). 目标变量熵: $H(D) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} \approx 0.811$.

"是否认真学习"的信息增益率:

$H(D|认真学习) = \frac{5}{8}\times 0 + \frac{3}{8}\times 0.918 \approx 0.344$

信息增益率: $GR_1 = \frac{H(D)-H(D|认真学习)}{-\frac{5}{8}\log_2\frac{5}{8}-\frac{3}{8}\log_2\frac{3}{8}} \approx 0.489$.

"是否超常发挥"信息增益率:

$H(D|超常发挥) = \frac{1}{2}\times 0.811 + \frac{4}{8}\times 0.811 = 0.811$

信息增益率: $G(R)_2 = \frac{0}{1} = 0$

∴"是否认真学习"更佳.

是否认真学习

是 通过

否 不通过

2).

是否认真学习

是 通过

否 是否超常发挥

是 通过

否 不通过

决策结果与训练数据一致