一、（20 分）对以下数据一维线性回归 y=wx+b

X	0	2	4
Y	1	3	7

请列出平方损失函数L，并直接通过令$\frac{\partial L}{\partial w}=0, \frac{\partial L}{\partial b}=0$，求出最小化$L$时 w,b 的数值解。请画出得到的回归曲线。

$$
\begin{aligned}
L &= \sum (y_i - wx_i - b)^2 \\
&= (1-b)^2 + (3-2w-b)^2 + (7-4w-b)^2 \\
&= 59 + 20w^2 + 3b^2 + 12wb - 68w - 22b
\end{aligned}
$$

$$
\begin{cases}
\frac{\partial L}{\partial w} = 2(3-2w-b)\times(-2) + 2(7-4w-b)\times(-4) = 0 \\
\frac{\partial L}{\partial b} = -2(1-b) - 2(3-2w-b) - 2(7-4w-b) = 0
\end{cases}
$$

$\Rightarrow w = \frac{3}{2}, b = \frac{2}{3}$ 故 $y = \frac{3}{2}x + \frac{2}{3}$

二、（20 分）课上学的逻辑回归以{1, -1}作为正负类标签，本题使用{1, 0}作为正负类标签。给定数据集$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$。设权重 (weight)为$w \in \mathbb{R}^d$ 和 偏置 (bias) 为 $b \in \mathbb{R}$，σ表示 sigmoid 函数

1)（6 分）写出$p(y = y_i | x = x_i)$在$y_i = 0, 1$下分别是多少。

2)（14 分）利用$p(y = y_i | x = x_i) = p(y = 1 | x = x_i)^{y_i}\, p(y = 0 | x = x_i)^{1-y_i}$，推导逻辑回归在 D 上的对数似然函数（log-likelihood）。

1) $p(y=1 | x=x_i) = \sigma(f(x)) = \sigma(wx_i + b)$

$p(y=0 | x=x_i) = 1 - \sigma(f(x)) = \sigma(-f(x)) = \sigma(-wx_i - b)$

故 $p(y=y_i | x=x_i) = \sigma((-1)^{y_i+1} f(x_i))$

2)
$$
\begin{aligned}
L &= \log \prod_{i=1}^{n} p(y=y_i | x=x_i) \\
&= \sum_{i=1}^{n} \log p(y=y_i | x=x_i) = \sum_{i=1}^{n} \left(y_i \log p(y=1 | x=x_i) + (1-y_i) \log p(y=0 | x=x_i) \right) \\
&= \sum_{i=1}^{n} \left(y_i \log \sigma(f(x)) + (1-y_i) \log \sigma(-f(x)) \right) \\
&= \sum_{i=1}^{n} \left(y_i \log \sigma(wx_i + b) + (1-y_i) \log \sigma(-wx_i - b) \right) \\
&= -\sum_{i=1}^{n} \left(y_i \log(1 + e^{-(wx_i+b)}) + (1-y_i) \log(1 + e^{wx_i+b}) \right)
\end{aligned}
$$

三、（30 分）利用树模型对以下数据进行二分类。id 表示数据编号，A，B，C 是特征，y 是标签。

id	1	2	3	4	5	6	7	8	9
A	0	0	0	0	1	1	1	1	1
B	1	1	1	1	0	1	0	1	1
C	0	1	1	1	0	0	1	1	1
y	1	-1	-1	-1	1	-1	-1	1	1

1) （15 分）在树的根节点，特征 A 的信息增益率 (information gain ratio) 是多少？(请使用以 2 为底的对数)

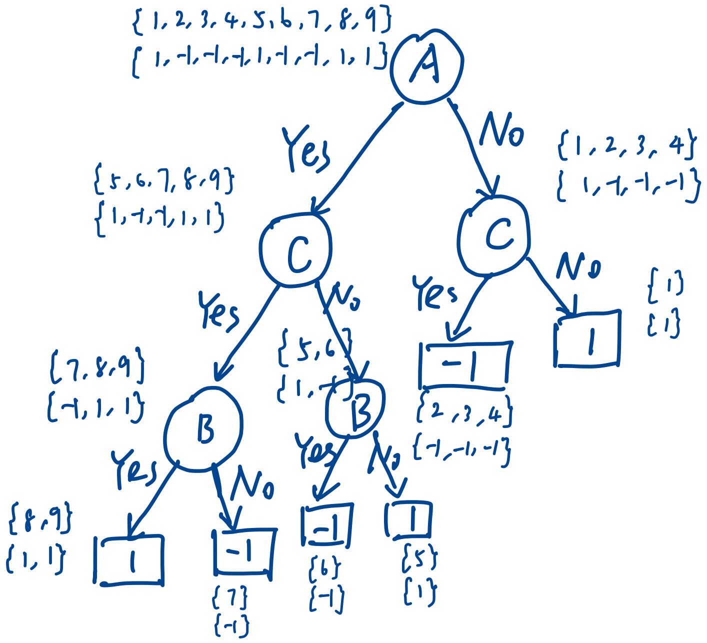
2) （15 分）假设在根节点对 A 分裂。在第二层所有结点对 C 分裂，在第三层对 B 分裂。请画出分类树并预测$x_* = [1,1,1]$的标签。

1) $$g(D,A) = H(D) - \sum \frac{|D^{A=a_i}|}{|D|} H(D^{A=a_i})$$

$$= -\frac{4}{9}\log\frac{4}{9} - \frac{5}{9}\log\frac{5}{9} - \left[\frac{4}{9}\times\left(-\frac{1}{4}\log\frac{1}{4}-\frac{3}{4}\log\frac{3}{4}\right) + \frac{5}{9}\times\left(-\frac{3}{5}\log\frac{3}{5}-\frac{2}{5}\log\frac{2}{5}\right)\right]$$

$$= 0.091$$

2)



$x_* = [1,1,1]$ 时，$y_* = 1$

四、（30 分）推导 softmax, log softmax 的反向传播公式。设输入$z \in \mathbb{R}^d$，计算图为线性（计算结点之间顺序连接，没有跨层连接），总损失函数为L。

1) （15 分）softmax 的输出为$a \in \mathbb{R}^d, a_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$。用$\frac{\partial L}{\partial a}$来表示$\frac{\partial L}{\partial z}$

2) （15 分）log softmax 的输出为$a \in \mathbb{R}^d, a_i = \ln\frac{e^{z_i}}{\sum_j e^{z_j}}$。用$\frac{\partial L}{\partial a}$来表示$\frac{\partial L}{\partial z}$

提示：逐分量表示$\frac{\partial L}{\partial z_i}$。先求$\frac{\partial a_j}{\partial z_i}$，再利用使用链式法则$\frac{\partial L}{\partial z_i} = \sum_j \frac{\partial L}{\partial a_j}\frac{\partial a_j}{\partial z_i}$。你可以使用$a$来表示$\frac{\partial L}{\partial z}$，最终表达式中不要出现 z。

1) $$\frac{\partial a_j}{\partial z_i} = \begin{cases} -a_i a_j, & i \neq j \\ a_i - a_i^2, & i = j \end{cases}$$

$$\frac{\partial L}{\partial z_i} = \sum_j \frac{\partial L}{\partial a_j}\cdot\frac{\partial a_j}{\partial z_i} = \sum_{j\neq i}\frac{\partial L}{\partial a_j}(-a_i a_j) + \frac{\partial L}{\partial a_i}(a_i - a_i^2)$$

$$= \sum_j \frac{\partial L}{\partial a_j}(-a_i a_j) + \frac{\partial L}{\partial a_i}\cdot a_i$$

$$= a_i\left(\frac{\partial L}{\partial a_i} - \sum_j a_j \frac{\partial L}{\partial a_j}\right)$$

$$\frac{\partial L}{\partial z} = \begin{pmatrix} \frac{\partial L}{\partial z_1} \\ \vdots \\ \frac{\partial L}{\partial z_d} \end{pmatrix} = \begin{pmatrix} a_1\frac{\partial L}{\partial a_1} \\ \vdots \\ a_d\frac{\partial L}{\partial a_d} \end{pmatrix} - \left(\sum_j a_j \frac{\partial L}{\partial a_j}\right)\cdot a$$

$$= a * \frac{\partial L}{\partial a} - \left(a^T\cdot\frac{\partial L}{\partial a}\right) a$$

2) $$\frac{\partial a_j}{\partial z_i} = \begin{cases} -a_i \;, & i \neq j \\ 1-a_i \;, & i = j \end{cases}$$

$$\frac{\partial L}{\partial z_i} = \sum_j \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_i} = \sum_{j \neq i} \frac{\partial L}{\partial a_j} \cdot (-a_i) + \frac{\partial L}{\partial a_i}(1-a_i)$$

$$= \sum_j \frac{\partial L}{\partial a_j}(-a_i) + \frac{\partial L}{\partial a_i}$$

$$\frac{\partial L}{\partial z} = \begin{pmatrix} \frac{\partial L}{\partial z_1} \\ \vdots \\ \frac{\partial L}{\partial z_d} \end{pmatrix} = \frac{\partial L}{\partial a} - a \cdot \sum_j \frac{\partial L}{\partial a_j} = \frac{\partial L}{\partial a} - \mathrm{sum}\left(\frac{\partial L}{\partial a}\right) \cdot a$$