1. (1)

   $W=(XX^T)^{-1}XY^T=[1,0,1]^T$

   (2)

   $XX^T=\begin{bmatrix}13 & 16 & 5 & 32\\ 16 & 21 & 7 & 42\\ 5 & 7 & 3 & 14\\ 32 & 42 & 14 & 84\end{bmatrix}$, rank($XX^T$) = 3 < 4. 该矩阵退化，因此解不唯一。

   (3)

   $W=(XX^T+\lambda I)^{-1}XY^T=[0.37798, 0.13988, 0.27976, 0, 30357]^T$

Note: 本题（以及下一题）应当注意X（以及Y, W等）的shape。本参考答案中的$X\in\mathbb{R}^{d\times n}$，课上讲解中$X\in\mathbb{R}^{n\times d}$（默认情况），若没有特殊注明，会扣除一小部分分数。

---

2. (1)

   $L=(W^TX-Y)R(W^TX-Y)^T+\lambda W^TW$
   $\frac{\partial L}{\partial W}=2\lambda W+2XR(W^TX-Y)^T$
   令 $\frac{\partial L}{\partial W}=0$ -> $W=(XRX^T+\lambda I)^{-1}XRY^T$

   (2)

   具有更高频率的数据点应当在加权回归中被赋予更高的权重。

   这两种情况是等价的。

---

3. (1)

   $L=-\sum t_i\log f(x_i)+(1-t_i)\log(1-f(x_i))$

   (2)

   t_i 可以被看作是$(x_i,1)$出现的次数，$1-t_i$可以被看作是$(x_i,0)$出现的次数。

   在这种情况下，likelihood是：$Likelihood=\prod_i^n P(y=1\mid x)^{t_i}\prod_i^n P(y=0\mid x)^{1-t_i}$.

   因此可以验证，$\log Likelihood=L$.

Note: 第2问应当从最大似然框架出发，写出并解释soft label所对应的Likelihood，进而也就解释了为什么L具有这样的形式。

---

4. (1)

   令 $f'(x)=sigmoid(f(x))=sigmoid(w^Tx+b)=P(y=1\mid x)=1-P(y=0\mid x)$
   $Likelyhood=\Pi_iP(y=y_i\mid x_i)$

因为 $w^Tx+b=0$ 是一个分割面，因此对于任意i：

1. $y_i=1 \Rightarrow f(x_i)>0$
2. $y_i=0 \Rightarrow f(x_i)<0$

当scale w 与 b 使得 $||kw|| \to +\infty$时，有：

1. $y_i=1 \Rightarrow f(x_i) \to +\infty \Rightarrow f'(x) \to 1$
2. $y_i=0 \Rightarrow f(x_i) \to -\infty \Rightarrow f'(x) \to 0$

因此 $P(y=y_i \mid x_i) \to 1$

即 $Likelyhood \to 1$

Note: 最大似然解要求似然概率likelihood取到最大值。部分同学仅仅解释了(kw^Tx+kb)是一个分割面，因此不正确。

(2)

$L'=-\sum_i y_i \log f(x_i)+(1-y_i)\log(1-f(x_i))+\lambda w^Tw$

- 非无穷

当 $||w|| \to \infty$，L' 同样趋近于无穷 (λw^Tw 增长速度比 $-\sum_i y_i \log f(x_i)+(1-y_i)\log(1-f(x_i))$快).

- 唯一性

$\frac{\partial L'}{\partial w}=-\sum_i \left(y_i-P(y=1 \mid x_i)\right)x_i+2\lambda w$
$\frac{\partial^2 L'}{\partial w \partial w^T}=\sum_i P(y=1|x_i)P(y=0|x_i)x_ix_i^T+2\lambda I$
由于 $\frac{\partial^2 L'}{\partial w \partial w^T}$ 处处正定, 因此L' 是一个严格凸函数, 即 L' 具有不大于1个全局最小值。同时由于当$||w|| \to \infty$时 $L' \to \infty$，因此L'具有唯一的全局最小值。

Note: 应当分别解释“不再无穷”和“具有唯一解”；应当分析出“**严格**凸函数”；同时注意，“严格凸函数”可能不存在全局最小值，例如e^x。因此，最完整的解释是：首先指出无穷远处L会爆炸，进一步分析严格凸函数，排除类似于e^x的情况。