1.a

Logistic loss function is

$$\begin{aligned} L(y, f(x)) &= log(1 + e^{-yf(x)}) \\ &= log(s^{-1}(yf(x))) \quad \text{where } y \in \{-1, 1\} \text{ and } s(\gamma) \text{ is the Sigmoid function} \end{aligned}$$

Logistic loss minimization with $Y \in \{-1, 1\}$ is

$$\begin{aligned} \min_f \ L(y, f(x)) &= \min_f \ log(1 + e^{-yf(x)}) \\ &= \min_f \ log\Big(s^{-1}(yf(x))\Big) \\ &= \max_f \ log\Big(s(yf(x))\Big) \end{aligned}$$

So we get,

$$\min_f \ L(y, f(x)) = \max_f \ log\Big(\frac{1}{1 + e^{-yf(x)}}\Big) \tag{0.0.1}$$

Using Equation 0.0.1, expected loss minimization is

$$\begin{aligned} \min_f \mathbb{E}[L(Y, f(X))] &= \min_f \ \mathbb{E}_X[\mathbb{E}_{Y|X}[L(Y, f(X))|X]] \\ &= \min_f \ \mathbb{E}_X\Big[\mathbb{P}(Y = 1|X)L(1, f(X)) + \mathbb{P}(Y = -1|X)L(-1, f(X))\Big] \\ &= \min_f \Big[\mathbb{P}(Y = 1|X)log(1 + e^{-f(X)}) + \mathbb{P}(Y = -1|X)log(1 + e^{f(X)})\Big] \\ &= \max_f \Big[\mathbb{P}(Y = 1|X)log\Big(\frac{1}{1 + e^{-f(X)}}\Big) + \mathbb{P}(Y = -1|X)log\Big(\frac{1}{1 + e^{f(X)}}\Big)\Big] \end{aligned}$$

So for each $X = x$ we get,

$$\min_f \mathbb{E}[L(Y, f(X))|X = x] = \max_f \Big[\mathbb{P}(Y = 1|x)log\Big(\frac{1}{1 + e^{-f(x)}}\Big) + \mathbb{P}(Y = -1|x)log\Big(\frac{1}{1 + e^{f(x)}}\Big)\Big] \tag{0.0.2}$$

With likelihood maximization where $Y \in \{0, 1\}$, we have:

$$L(y^n; p_1, ..., p_n) = \prod_i^n p^{y_i}(1 - p_i)^{1-y_i}$$

Maximization of Log Likelihood is

$$\begin{aligned}
\max_{\vec{\beta}} \ell(y; x, \vec{\beta}) &= \max_{\vec{\beta}} \sum_{i}^{n} \left[y_i log\Big(\frac{1}{1+e^{-x^T\vec{\beta}}}\Big) + (1-y_i) log\Big(\frac{1}{1+e^{x^T\vec{\beta}}}\Big)\right] \\
&= \max_{\vec{\beta}} \left[\sum_{i:y_i=1}^{n} log\Big(\frac{1}{1+e^{-x_i^T\vec{\beta}}}\Big) + \sum_{i:y_i=-1}^{n} log\Big(\frac{1}{1+e^{x_i^T\vec{\beta}}})\Big)\right] \\
&= \max_{\vec{\beta}}\ n\left[\frac{n_1}{n} log\Big(\frac{1}{1+e^{-x_i^T\vec{\beta}}}\Big) + \frac{n_0}{n} log\Big(\frac{1}{1+e^{x_i^T\vec{\beta}}})\Big)\right] \\
&= \max_{\vec{\beta}}\ n\left[\hat{\pi}_1 log\Big(\frac{1}{1+e^{-x_i^T\vec{\beta}}}\Big) + \hat{\pi}_0 log\Big(\frac{1}{1+e^{x_i^T\vec{\beta}}})\Big)\right]
\end{aligned}$$

So we have,

$$\max_{\vec{\beta}} \ell(y; x, \vec{\beta}) = \max_{\vec{\beta}}\ \left[\hat{\pi}_1 log\Big(\frac{1}{1+e^{-x_i^T\vec{\beta}}}\Big) + \hat{\pi}_0 log\Big(\frac{1}{1+e^{x_i^T\vec{\beta}}})\Big)\right] \tag{0.0.3}$$

Equation 0.0.2 and Equation 0.0.3 are equivalent if $\mathbb{P}(Y = k|x)$ are approximated by $\hat{\pi}_k = \frac{n_k}{n}$.

1.b

(a) Minimizer for the logistic loss function

We want to minimize the expected loss

$$
\begin{aligned}
\min_f \mathbb{E}[L(Y, f(X))] &= \min_f \ \mathbb{E}_X[\mathbb{E}_{Y|X}[L(Y, f(X))|X]] \\
&= \min_f \ \mathbb{E}_X\Big[\mathbb{P}(Y=1|X)L(1, f(X)) + \mathbb{P}(Y=-1|X)L(-1, f(X))\Big]
\end{aligned}
$$

For each $X = x$, we want to minimize

$$
\mathbb{E}[L(Y, f(X))|X = x] = \mathbb{P}(Y=1|X=x)L(1, f(x)) + \mathbb{P}(Y=-1|X=x)L(-1, f(x))
$$

Take derivative with respect to f and set to 0

$$
\frac{\mathrm{d}}{\mathrm{d}f}\mathbb{P}(Y=1|x)log(1+e^{-f(x)}) + \mathbb{P}(Y=-1|x)log(1+e^{f(x)})
$$

$$
\mathbb{P}(Y=1|X=x)\frac{1}{1+e^{-f(x)}}(-e^{-f(x)}) + \mathbb{P}(Y=-1|X=x)\frac{1}{1+e^{-f(x)}}e^{f(x)} = 0
$$

$$
\frac{\mathbb{P}(Y=1|X=x)}{\mathbb{P}(Y=-1|X=x)} = \frac{\frac{e^{f(x)}}{1+e^{f(x)}}}{\frac{e^{-f(x)}}{1+e^{-f(x)}}} = e^{f(x)}
$$

So the optimal function for logistic loss function is

$$
f^*(x) = log\left(\frac{\mathbb{P}(Y=1|X=x)}{\mathbb{P}(Y=-1|X=x)}\right) = log\left(\frac{\mathbb{P}(Y=1|X=x)}{1-\mathbb{P}(Y=1|X=x)}\right)
$$

(b) Minimizer for the hinge loss function

For each $X = x$, we want to minimize

$$
S = \mathbb{E}[L(Y, f(X))|X = x] = ((1+f(x))_+\mathbb{P}(Y=1|X=x) + ((1-f(x)))_+\mathbb{P}(Y=-1|X=x)
$$

If $f(x) \geqslant 1$, $S = (1+f(x))\mathbb{P}(Y=1|X=x)$ and S is minimized at its lower bound, which is $f(x) = 1$. So every point will be classified as 1.

If $f(x) \leqslant -1$, $S = (1-f(x))\mathbb{P}(Y=-1|X=x)$ and S is minimized at its upper bound, which is $f(x) = -1$. So every point will be classified as -1.

For $-1 \leqslant f(x) \leqslant 1$,

$$
\begin{aligned}
S = \mathbb{E}[L(Y, f(X))|X = x] &= (1-f(x)\mathbb{P}(Y=1|X=x) + (1+f(x))\mathbb{P}(Y=-1|X=x) \\
&= (1-f(x))(1-\mathbb{P}(Y=-1|X=x)) + (1+f(x))\mathbb{P}(Y=-1|X=x) \\
&= 1 - f(x)(1-2\mathbb{P}(Y=-1|X=x))
\end{aligned}
$$

$$
f^*(x) = \begin{cases} 1, & \text{if } \mathbb{P}(Y=1|X=x) \geqslant \mathbb{P}(Y=-1|X=x) \\ -1, & \text{if } \mathbb{P}(Y=1|X=x) < \mathbb{P}(Y=-1|X=x) \end{cases}
$$