Question 2, a.

Let $\theta \in R^d$, $x \in R^d$.

Model : $y_i = x_i^t \theta + \varepsilon_i$ w/ $\varepsilon_i \sim \text{Laplace}(0, b)$.

Data : $D = \{(x_i, y_i)\}_{i=1,2 \dots n}$.

Estimate by maximum likelihood

pdf of ε_i is $\frac{1}{2b} e^{-\frac{|z|}{b}}$ over $z \in R$.

likelihood of data using $\varepsilon_i = y_i - x_i^t \theta$.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2b} e^{-\frac{|y_i - x_i^t \theta|}{b}}$$ by independence of data.

$$\log L(\theta) = -\sum_{i=1}^{n} \frac{|y_i - x_i^t \theta|}{2b^2}$$ + terms constant in θ

maximizing data likelihood $\Leftrightarrow$

minimizing $$\sum_{i=1}^{n} \frac{|y_i - x_i^t \theta|}{2b^2}$$

i.e. $\min_{\theta} \frac{1}{2b^2} \sum_{i=1}^{n} |y_i - x_i^t \theta|$.

$\iff \min_{\theta} \sum_{i=1}^{n} |y_i - x_i^t \theta|$ as $2b^2 > 0$. $\frac{1}{2b^2} > 0$.

Above shown the maximum likelihood estimate of θ is given by $\min_{\theta} \sum_{i=1}^{n} |y_i - x_i^t \theta|$.

The loss function is

$$l(\hat{y}, y) = |\hat{y} - y|.$$

while we can still use square loss $|\hat{y} - y|^2$ as loss function because $|a| = |b| \iff a^2 = b^2$.

And square loss is more convenient for differentiation.

Question 2, b.

Estimate θ by maximum likelihood.

$$L(\theta) = \prod_{i=1}^{n} p(y = y_i \mid x^{(i)}; \theta)$$

$$p(y = y_i \mid x^{(i)}; \theta)$$
$$= f(x^{(i)}; \theta)_{y_i}$$
$$= f(x^{(i)}; \theta)_1^0 \, f(x^{(i)}; \theta)_2^0 \cdots f(x^{(i)}; \theta)_{y_i}^1 \cdots \quad (*)$$

As $y_i \in \{1, 2 \cdots K\}$.

$$(*) = \prod_{k=1}^{K} f(x^{(i)}; \theta)_k^{1_{y_i=k}}$$

$$\therefore L(\theta) = \prod_{i=1}^{n} \prod_{k=1}^{K} f(x^{(i)}; \theta)_k^{1_{y_i=k}}.$$

$$\log L(\theta) = \sum_{i=1}^{n} \log \prod_{k=1}^{K} f(x^{(i)}; \theta)_k^{1_{y_i=k}}$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} 1_{y_i=k} \log f(x^{(i)}; \theta)_k$$

to maximize $\log L(\theta)$ is to minimize

$$-\sum_{i=1}^{n} \sum_{k=1}^{K} 1_{y_i=k} \log f(x^{(i)}; \theta)_k.$$

Q.E.D.

Question 3, a

$$\sigma(z) = \frac{e^z}{e^z+1} = 1 - \frac{1}{e^z+1}.$$

$$= 1 - (e^z+1)^{-1}.$$

$$\sigma'(z) = (e^z+1)^{-2} \cdot (e^z+1)'$$

$$= (e^z+1)^{-2} \cdot e^z$$

$$= \frac{e^z}{(e^z+1)^2}$$

$$= \frac{e^z}{e^z+1} \cdot \frac{1}{e^z+1}$$

$$= \sigma(z)(1-\sigma(z))$$

Q.E.D.

Question 3, b

$$f(\theta) = \sum_{i=1}^{n} -y_i \log \hat{y}_i - (1-y_i) \log (1-\hat{y}_i)$$

$$\hat{y}_i = \sigma(\theta x_i) . \qquad \sigma'(z) = \sigma(z)(1-\sigma(z)) .$$

$$\frac{\partial f}{\partial \theta_j} = \sum_{i=1}^{n} (-y_i \log \hat{y}_i - (1-y_i) \log (1-\hat{y}_i))$$

$$= \sum_{i=1}^{n} \left(-y_i \frac{\partial \log \hat{y}_i}{\partial \theta_j} - (1-y_i) \frac{\partial \log (1-\hat{y}_i)}{\partial \theta_j}\right)$$

$$= \sum_{i=1}^{n} \left(-y_i \frac{\partial \log \hat{y}_i}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial \theta_j} - (1-y_i) \frac{\partial \log (1-\hat{y}_i)}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial \theta_j}\right)$$

$$= \sum_{i=1}^{n} \cdot \frac{\partial \hat{y}_i}{\partial \theta_j} \left(-y_i \frac{1}{\hat{y}_i} - (1-y_i) \frac{-1}{1-\hat{y}_i}\right)$$

$$= \sum_{i=1}^{n} \frac{\partial \hat{y}_i}{\partial \theta_j} \cdot \frac{\hat{y}_i - y_i}{\hat{y}_i (1-\hat{y}_i)}$$

$$\frac{\partial \hat{y}_i}{\partial \theta_j} = \frac{\partial \sigma(\theta x_i)}{\partial \theta_j} = \frac{\partial \sigma(\theta x_i)}{\partial \theta x_i} \cdot \frac{\partial \theta x_i}{\partial \theta_j}$$

$$= \sigma(\theta x_i)(1-\sigma(\theta x_i)) \cdot \frac{\partial \sum_{j=1}^{d} \theta_j x_{ij}}{\partial \theta_j}$$

$$= \hat{y}_i(1-\hat{y}_i) \cdot x_{ij}$$

$$\therefore \frac{\partial f}{\partial \theta_j} = \sum_{i=1}^{n} \hat{y}_i(1-\hat{y}_i) \cdot x_{ij} \cdot \frac{\hat{y}_i - y_i}{\hat{y}_i(1-\hat{y}_i)}$$

$$= \sum_{i=1}^{n} x_{ij}(\hat{y}_i - y_i).$$

$$\therefore \frac{\partial f}{\partial \theta} = \sum_{i=1}^{n} x_i(\hat{y}_i - y_i).$$

$$= x \cdot (\hat{y} - y)$$