$$\hat{p} = \frac{\exp(w^T x)}{1+\exp(w^T x)} = \frac{1}{1+\exp(-w^T x)} = \hat{y}$$

$$L(w) = -[y \cdot \log(\hat{y}) + (1-y) \cdot \log(1-\hat{y})]$$

$$\frac{\partial L(w)}{\partial w} = \frac{\partial L(w)}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial L(w)}{\partial \hat{y}} = -\left(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}\right)$$

$$= \frac{\hat{y} - y}{\hat{y}(1-y)}$$

$$\frac{\partial z}{\partial w} = X$$

$$\frac{\partial \hat{y}}{\partial z} = \frac{\partial}{\partial z}\left[\frac{1}{1+\exp(w^T x)}\right]$$

$$= \frac{\partial}{\partial z}(1+\exp(w^T x))^{-1}$$

$$= -(1+e^{w^T x})^{-2} \cdot (-e^{w^T x})$$

$$= \frac{e^{-w^T x}}{(1+e^{-w^T x})^2}$$

$$= \frac{1}{(1+e^{-w^T x})} \cdot \frac{e^{-w^T x}}{(1+e^{-w^T x})}$$

$$= \frac{1}{1+e^{-w^T x}} \cdot \frac{(1+e^{-w^T x})-1}{1+e^{-w^T x}}$$

$$= \frac{1}{1+e^{-w^T x}} \cdot \left(\frac{1+e^{-w^T x}}{1+e^{-w^T x}} - \frac{1}{1+e^{-w^T x}}\right)$$

$$= \frac{1}{1+e^{-w^T x}} \cdot \left(1 - \frac{1}{1+e^{-w^T x}}\right)$$

$$= \hat{y}(1-\hat{y})$$

$$\frac{\partial L(w)}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w} = X \cdot \left(\frac{\hat{y}-y}{\cancel{\hat{y}(1-\hat{y})}}\right) \cdot \cancel{\hat{y}(1-\hat{y})}$$

$$\boxed{\text{gradient} = X(\hat{y}-y)}$$