Logistic Loss Gradient:

$$L(\beta) = -\sum_{i=1}^{n} y_i \log\Big(\overbrace{\frac{1}{1+e^{-\langle \beta, x_i\rangle}}}^{h_\beta(x)}\Big) + (1-y_i)\log\Big(\overbrace{1-\frac{1}{1+e^{-\langle \beta, x_i\rangle}}}^{1-h_\beta(x)}\Big)$$

Note that $\dfrac{1}{1+e^{-\langle \beta, x_i\rangle}} = \dfrac{e^{\langle \beta, x_i\rangle}}{e^{\langle \beta, x_i\rangle}+1}$ so $1-\dfrac{1}{1+e^{-\langle \beta, x_i\rangle}} = \dfrac{1}{1+e^{\langle \beta, x_i\rangle}}$

So, $L(\beta) = \sum_{i=1}^{n} y_i \log(1+e^{-\langle \beta, x_i\rangle}) + (1-y_i)\log(1+e^{\langle \beta, x_i\rangle})$

Computing each partial seperately:

$$\frac{\partial}{\partial \beta_i} L(\beta) = \sum_{i=1}^{n} y_i \frac{\partial}{\partial \beta_i}\log(1+e^{-\langle \beta, x_i\rangle}) + (1-y_i)\frac{\partial}{\partial \beta_i}\log(1+e^{\langle \beta, x_i\rangle})$$

(by chain rule) $$= \sum_{i=1}^{n} y_i \left[\frac{\partial}{\partial \beta_i} e^{-\langle \beta, x_i\rangle}\right]\frac{1}{1+e^{-\langle \beta, x\rangle}} + (1-y_i)\left[\frac{\partial}{\partial \beta_i} e^{\langle \beta, x_i\rangle}\right]\frac{1}{1+e^{\langle \beta, x\rangle}}$$

(chain rule again) $$= \sum_{i=1}^{n} -y_i x_i \frac{e^{-\langle \beta, x_i\rangle}}{1+e^{-\langle \beta, x_i\rangle}} + (1-y_i)x_i \frac{e^{\langle \beta, x_i\rangle}}{1+e^{\langle \beta, x_i\rangle}}$$

$$= \sum_{i=1}^{n} -y_i x_i (1-h_\beta(x_i)) + (1-y_i)x_i h_\beta(x_i)$$

$$= \sum_{i=1}^{n} x_i (h_\beta(x_i) - y_i)$$

$$\nabla L(\beta) = \begin{bmatrix} \partial/\partial\beta_1 \\ \vdots \\ \partial/\partial\beta_d \end{bmatrix} = X^T(h_\beta(X) - y_i)$$

where $h_\beta(X) = \begin{bmatrix} h_\beta(x_1) \\ \vdots \\ h_\beta(x_n) \end{bmatrix}$