1 (a) Note that $\frac{1}{1+e^z} = 1 - \sigma(z)$ and $\frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1-\sigma(z))$.

$$\frac{\partial E(w)}{\partial w_j} = \sum_{i=1}^{N} (x_j)^{(i)} y^{(i)} (1 - \sigma(w^T x^{(i)})) - (x_j)^{(i)} \sigma(w^T x^{(i)}) + (x_j)^{(i)} y^{(i)} \sigma(w^T x^{(i)})$$

$$= \sum_{i=1}^{N} (x_j)^{(i)} y^{(i)} - (x_j)^{(i)} \sigma(w^T x^{(i)})$$

$$\frac{\partial^2 E(w)}{\partial (w_j)^2} = -\sum_{i=1}^{N} (x_j)^{(i)} \sigma(w^T x^{(i)})(1 - \sigma(w^T x^{(i)}))(x_j)^{(i)}$$

$$\frac{\partial^2 E(w)}{\partial w_j w_k} = -\sum_{i=1}^{N} (x_j)^{(i)} \sigma(w^T x^{(i)})(1 - \sigma(w^T x^{(i)}))(x_k)^{(i)}$$

Let $X \in \mathbb{R}^{n \times m}$ be the design matrix and $w \in \mathbb{R}^m$ be the weight vector, where n is the number of observations and m is the number of features. We can express the second-order partial derivatives in matrix form as

$$X^T diag[\sigma(Xw)(1 - \sigma(Xw))]X$$

which gives us our Hessian matrix.