- The gradient of any function is defined as

$$\vec{\nabla} f(\mathbf{x}) = \sum_{i=1}^{\mathcal{D}} \frac{\partial f(\mathbf{x})}{\partial x_i} \hat{e}_i \tag{47}$$

where $\mathcal{D}$ is the dimensionality/number of dependent variables in $f(\mathbf{x})$ and $\hat{e}_i$ is the normal vector for that coordinate – a value of 1 for all dimensions in cartesian space. This requires a derivative over $\mathbf{x}$ and $\mathbf{w}$ which is

$$\vec{\nabla} f(\mathbf{x}_i) = \sum_{i=1}^{\mathcal{D}} \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{w}} + \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \tag{48}$$

We can rewrite our equation $f(\mathbf{x}_i)$ to make it easier to differentiate

$$f(\mathbf{x}_i) = \sum_{i=1}^{n} \log\left(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)\right) + \frac{||\mathbf{w}||^2}{\sigma^2} \tag{49}$$

where the derivative of a sum of a function is the sum of the derivative of that function, so the differentials can be brought inside the sum and the respected differentials can be taken and the partial derivative outside the braces denotes that segment representing that differential of $f(\mathbf{x}_i)$

$$\vec{\nabla} f(\mathbf{x}_i) = \left\{ \sum_{i=1}^{n} \frac{-y_i \mathbf{x}_i}{\exp(y_i \mathbf{w}^T \mathbf{x} + 1)} + \frac{2\,||\mathbf{w}||}{\sigma^2} \right\}_{\frac{\partial f}{\partial \mathbf{w}}} + \left\{ \sum_{i=1}^{n} \frac{-y_i \mathbf{w}^T}{\exp(y_i \mathbf{w}^T \mathbf{x}) + 1} \right\}_{\frac{\partial f}{\partial \mathbf{x}}} \tag{50}$$

2. [5 points] Find the Hessian of (a).

- The Hessian Matrix is defined as

$$\boldsymbol{\mathcal{H}}\left(f(\mathbf{x}_i)\right) = \begin{bmatrix} \frac{\partial^2 f}{\partial \mathbf{x}_i^2} & \frac{\partial^2 f}{\partial \mathbf{x}_i \partial \mathbf{w}} \\ \frac{\partial^2 f}{\partial \mathbf{w} \partial \mathbf{x}_i} & \frac{\partial^2 f}{\partial \mathbf{w}^2} \end{bmatrix} \tag{51}$$

$$\frac{\partial^2 f(\mathbf{x}_i)}{\partial \mathbf{x}_i^2} = \sum_{i=1}^{n} \frac{||\mathbf{w}||^2 y_i^2 e^{\varsigma_i}}{(e^{\varsigma_i} + 1)^2} \tag{52}$$

$$\frac{\partial^2 f(\mathbf{x}_i)}{\partial \mathbf{x}_i \partial \mathbf{w}} = \sum_{i=1}^{n} \frac{y_i \left[e^{\varsigma_i}\left(\varsigma_i - 1\right) - 1\right]}{(e^{\varsigma_i} + 1)^2} \tag{53}$$

$$\frac{\partial^2 f(\mathbf{x}_i)}{\partial \mathbf{w} \partial \mathbf{x}_i} = \frac{\partial^2 f(\mathbf{x}_i)}{\partial \mathbf{x}_i \partial \mathbf{w}} = \sum_{i=1}^{n} \frac{y_i \left[e^{\varsigma_i}\left(\varsigma_i - 1\right) - 1\right]}{(e^{\varsigma_i} + 1)^2} \tag{54}$$

$$\frac{\partial^2 f(\mathbf{x}_i)}{\partial \mathbf{w}^2} = \sum_{i=1}^{n} \frac{||\mathbf{x}_i||^2 y_i^2 e^{\varsigma_i}}{(e^{\varsigma_i} + 1)^2} + \frac{2}{\sigma^2} \tag{55}$$

where $\varsigma_i = \mathbf{w}^T \mathbf{x}_i y_i$