1a) Consider the log likelihood function for logistic regression

$l(\theta) = \sum_{i=1}^{m} y^i \log\ \mathrm{h}(x^i) + (1-y^i)\log(1-h(x^i))$ find the Hessian H of this function, and show that for any vextor z it holds true that $z^T Hz \leq 0$

To find the Hessian

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$l(\theta) = y^i log\big(g(\theta^T x)\big) + (1-y^i)\log(1-g(\theta^T x))$$

$$\frac{\partial l(\vartheta)}{\partial \theta_j} = y^i \frac{\frac{\partial g(\theta^T x)}{\partial \theta_j}}{g(\theta^T x)} + (1-y^i)\frac{\partial \log\big(1-g(\theta^T x)\big)}{\partial \theta_j} = y^i \frac{\frac{\partial g(\theta^T x)}{\partial \theta_j}}{g(\theta^T x)} + (1-y^i)\frac{\frac{\partial(1-g(\theta^T x))}{\partial \theta_j}}{1-g(\theta^T x)}$$

$$= y^i \frac{\frac{\partial g(\theta^T x)}{\partial \theta_j}}{g(\theta^T x)} - (1-y^i)\frac{\frac{\partial(g(\theta^T x))}{\partial \theta_j}}{1-g(\theta^T x)}$$

$$g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}; g(z) = \frac{1}{1+e^{-z}}; \frac{\partial g(z)}{\partial z} = \frac{\partial(1+e^{-z})^{-1}}{\partial z} = (-1)(1+e^{-z})^{-2}\frac{\partial(1+e^{-z})}{\partial z} = \frac{-1}{(1+e^{-z})^2}(-1)e^{-z}dz$$

$$= \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1-1+e^{-z}}{(1+e^{-z})^2} = \frac{1}{(1+e^{-z})}\left(1-\frac{1}{(1+e^{-z})}\right) = g(z)(1-g(z))$$

Now we can apply this to computing the Hessian:

$$y^i \frac{\frac{\partial g(\theta^T x)}{\partial \theta_j}}{g(\theta^T x)} - (1-y^i)\frac{\frac{\partial(g(\theta^T x))}{\partial \theta_j}}{1-g(\theta^T x)} = \left[\frac{y}{g(\theta^T x)} - \frac{(1-y)}{1-g(\theta^T x)}\right]\frac{\partial g(\theta^T x)}{\partial \theta_j}$$

$$= \left[\frac{y}{g(\theta^T x)} - \frac{(1-y)}{1-g(\theta^T x)}\right] g(\theta^T x)(1-g(\theta^T x))\frac{\partial(\theta^T x)}{\partial \theta_j}$$

$$= \left[\frac{y}{g(\theta^T x)} g(\theta^T x) - \frac{(1-y)}{1-g(\theta^T x)} g(\theta^T x)\right]\big(1-g(\theta^T x)\big)\frac{\partial(\theta^T x)}{\partial \theta_j}$$

$$= \left[y - \frac{(1-y)}{1-g(\theta^T x)} g(\theta^T x)\right]\big(1-g(\theta^T x)\big)\frac{\partial(\theta^T x)}{\partial \theta_j}$$

$$= \left[y\left(1 - g(\theta^T x)\right) - (1 - y)g(\theta^T x)\right] \frac{\partial(\theta^T x)}{\partial\theta_j} = \left[\left(y - yg(\theta^T x)\right) - \left(g(\theta^T x) - yg(\theta^T x)\right)\right] \frac{\partial(\theta^T x)}{\partial\theta_j}$$

$$= [y - g(\theta^T x)] \frac{\partial(\theta^T x)}{\partial\theta_j} = [y - h_\vartheta(x)]x_j$$

$$\frac{\partial^2 y}{\partial\theta_k} = \frac{\partial[y - h_\theta(x))x_j]}{\partial\theta_k} = x_j \frac{\partial[-h_\theta(x))]}{\partial\theta_k} = -x_j \frac{\partial g(\theta^T x)}{\partial\theta_k} = -x_j g(z)(1 - g(z)) \frac{\partial z}{\partial\theta_k}$$

$$-x_j h_\theta(x)\left(1 - h_\theta(x)\right)x_k = Xh_\theta(x)(1 - h_\theta(x))X^T$$