Classification_extra_credit

$$\frac{\partial f(x,w)}{\partial w}, where f(x,w) = \frac{1}{1+e^{-w^t x}}$$

Using the chain rule and assuming that $z = e^{-w^t x}$ we have:

$$h(z) = \frac{1}{1+e^z} \frac{\partial h(z)}{\partial w} = \frac{\partial h(z)}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial h(z)}{\partial z} = -\frac{1}{(1+e^z)^2} \cdot e^z = -\frac{e^z}{(1+e^z)^2}$$

$$\frac{\partial z}{\partial w} = -x \cdot e^{-w^t x}$$

$$\frac{\partial f(x,w)}{\partial w} = -\frac{e^{-w^t x}}{(1+e^{-w^t x})^2} \cdot (-x) = \frac{x e^{-w^t x}}{(1+e^{-w^t x})^2}$$

Now, the gradient of l(w) that is negative log likelihood is:

$$\nabla l(w) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{y_i - f(x_i,w)}{f(x_i,w)(1-f(x_i,w))} \right) x_i$$