HW 4

5.1.) $f(x) = \log(x^4)\sin(x^3) = 4\log(x)\sin(x^3)$

$$\boxed{f'(x) = \frac{4\sin(x^3)}{x} + 3x^2\log(x^4)\cos(x^3)}$$

5.2.) $f(x) = \frac{1}{1+e^{-x}}$

$$f'(x) = \frac{\frac{d}{dx}(1)(1+e^{-x}) - \frac{d}{dx}(1+e^{-x})}{(1+e^{-x})^2} \Rightarrow \boxed{f'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{x}}{(e^{x}+1)^2}}$$

3.) $f(x,y,z) = xe^y - x + z^3 - z$

a.) $\vec{\nabla} f(x,y,z) = (e^y - 1,\ xe^y,\ 3z^2 - 1)$

b.) $f_x = e^y - 1 = 0 \Rightarrow y = 0$

$f_y = xe^y = 0 \Rightarrow x = 0$

$f_z = 3z^2 - 1 = 0 \Rightarrow z = \pm\frac{1}{\sqrt{3}}$

$\boxed{(0, 0, \pm 1/\sqrt{3})}$

c.) $f(1.1, 0.1, 1) \approx f(1,0,1) + \overset{0}{\cancel{\frac{\partial f}{\partial x}(1,0,1)}} \cdot 0.1 + \overset{1}{\cancel{\frac{\partial f}{\partial y}(1,0,1)}} \cdot 0.1 = 0 + 0.1 = 0.1$

$f(1.1, 0.1, 1) = 0.115688$ ACTUAL

4.) $\vec{f}(\vec{x}) = \begin{bmatrix} \cos(x_1x_2) \\ \sin(x_2x_4) \\ x_1^2 + x_3^2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

a.) $$\frac{d\vec{f}}{d\vec{x}}(\vec{x}) = \begin{bmatrix} -x_2\sin(x_1x_2) & -x_1\sin(x_1x_2) & 0 & 0 \\ 0 & x_4\cos(x_2x_4) & 0 & x_2\cos(x_2x_4) \\ 2x_1 & 0 & 2x_3 & 0 \end{bmatrix}$$

b.) $\vec{f}(1, 0, 2.1, -0.1) = \vec{f}((1,0,2,0) + (0,0,.1,-.1))$

$$\approx \vec{f}(1,0,2,0) + \frac{d\vec{f}}{d\vec{x}}(1,0,2,0)\begin{bmatrix} 0 \\ 0 \\ .1 \\ -.1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 4 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ .1 \\ -.1 \end{bmatrix} = \boxed{\begin{bmatrix} 1 \\ 0 \\ 5.4 \end{bmatrix} \approx \vec{f}(1,0,2.1,-0.1)}$$

$$\boxed{\text{Actual: } \vec{f}(1,0,2.1,-0.1) = \begin{bmatrix} 1 \\ 0 \\ 5.41 \end{bmatrix}}$$

5.) $L(\vec{\theta}) = -\sum_{i=1}^{N}\left[y_i \log(g_{\vec{\theta}}(\vec{x}_i)) + (1-y_i)\log(1-g_{\vec{\theta}}(\vec{x}_i))\right]$

$g_{\vec{\theta}}(\vec{x}) = \frac{1}{1+e^{-\theta^T x}}$, where $\vec{\theta} = [\beta_0, \beta_1, \cdots, \beta_d]^T \in \mathbb{R}^{d+1}$

$\vec{\nabla} L(\vec{\theta}) = \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}, \cdots, \frac{\partial L}{\partial \beta_d}\right]$

$\frac{\partial L}{\partial \beta_j} = \frac{\partial L}{\partial g}\frac{\partial g}{\partial \beta_j} = \left(\frac{1-y_i}{1-g} - \frac{y_i}{g}\right)\left(x_j g(1-g)\right) = x_j g(1-y_i) - x_j y_i(1-g) = x_j g - x_j y_i g - x_j y_i + x_j y_i g$

$\Rightarrow \frac{\partial L}{\partial \beta_j} = x_j(g - y_i)$

$\frac{\partial L}{\partial g} = -\frac{y_i}{g} + \frac{1-y_i}{1-g}$

$\frac{\partial g}{\partial \beta_j} = x_j \frac{e^{-\theta^T x}(+1-1)}{(1+e^{-\theta^T x})^2} = x_j\left[\frac{1+e^{-\theta^T x}}{(1+e^{-\theta^T x})^2} - \frac{1}{(1+e^{-\theta^T x})^2}\right] = x_j\left[g_{\vec{\theta}}(\vec{x}) - g^2_{\vec{\theta}}(\vec{x})\right]$

$$\boxed{\therefore \vec{\nabla} L(\vec{\theta}) = \left[\sum_{i=1}^{N}(g_{\vec{\theta}}(x^{(i)}) - y^{(i)}), \sum_{i=1}^{N}(g_{\vec{\theta}}(x^{(i)}) - y^{(i)})x_1^{(i)}, \cdots, \sum_{i=1}^{N}(g_{\vec{\theta}}(x^{i}) - y^{i})x_N^{i}\right]}$$

6.) $\max(\nabla f(\vec{x}_0)\vec{v})$ when $\vec{v} = \frac{\nabla f(\vec{x}_0)^T}{\|\nabla f(\vec{x}_0)^T\|}$

$\vec{\nabla} f(\vec{x}_0)\vec{v} = (\vec{\nabla} f(\vec{x}_0)^T)^T\vec{v} = \langle \nabla f(\vec{x}_0)^T, \vec{v}\rangle \le \|\nabla f(x_0)^T\| \underbrace{\|\vec{v}\|}_{1 \text{ since } \vec{v} \text{ is unit}}$ by Cauchy-Schwarz

$\Rightarrow \langle \nabla f(\vec{x}_0)^T, \vec{v}\rangle = \|\nabla f(\vec{x}_0)^T\|$ iff $\vec{v}$ is parallel to $\nabla f(\vec{x}_0)^T$

$\therefore \vec{v} = \frac{\nabla f(\vec{x}_0)^T}{\|\nabla f(\vec{x}_0)^T\|}$ since $\vec{v}$ is unit & must be parallel to maximize QED