COEN 240

HW 1

Shiva Govindaraju

1 a) $f(x) = \frac{1}{2} x^T A x + b^T x$, A is symmetric.

$\nabla f(x) = \nabla (\frac{1}{2} x^T A x + b^T x)$

$= \frac{1}{2} \nabla (x^T A x) + \nabla (b^T x)$

$= \frac{1}{2} (2 A x) + (b^T)^T$

$$\boxed{\nabla f(x) = Ax + b}$$

b) $\nabla_x f(x) = \nabla_x (g(h(x)))$, let $u = h(x)$

$= \nabla_u (g(u)) \cdot \nabla_x u$, by chain-rule.

$$\therefore \boxed{\nabla_x f(x) = \nabla_{h(x)} g(h(x)) \cdot \nabla_x h(x)}$$

c) $f(x) = \frac{1}{2} x^T A x + b^T x$

$\nabla f(x) = \frac{\partial}{\partial x} f(x) = \frac{\partial}{\partial x} (\frac{1}{2} x^T A x + b^T x)$

$= Ax + b$ (answer to 1a)

$H = \nabla^2 f(x) = \frac{\partial}{\partial x} \nabla f(x) = \frac{\partial}{\partial x} (Ax + b) = \frac{\partial}{\partial x} Ax + \frac{\partial}{\partial x} b$

$$\boxed{H = A}$$

d) $f(x) = g(a^T x)$

$\nabla_x f(x) = \nabla_x g(a^T x)$ let $u = a^T x$

$= \nabla_u (g(u)) \cdot \nabla_x (a^T x)$

$= \nabla_u (g(u)) \cdot (a)$

$$\therefore \boxed{\nabla_x f(x) = a \cdot \nabla_u (g(u)) \text{ where } u = a^T x}$$

1 d, cont) $\nabla_x^2 f(x) = \nabla_{x_j} \nabla_{x_i} f(x)$

$= \nabla_{x_j} (a \cdot \nabla_{u_i} (g(u_i)))$ where $u_i = a^T x_i$

$= a \cdot \nabla_{u_j} \nabla_{u_i} (g(u_i)) \cdot \nabla_{x_j} (u_i)$ where $u_j = a^T x_j$

$$\boxed{\nabla_x^2 f(x) = a \cdot a^T \cdot \nabla_u^2 (g(u)) \quad \text{for } u = a^T x}$$

2 a) $A = zz^T$ where $z \in \mathbb{R}^n$

step 1: $(zz^T)^T = z^T (z^T)^T = z^T z$

$\therefore A = A^T \rightarrow A$ is a symmetric matrix.

step 2: $x^T A \cdot x = x^T (zz^T) x$

$= (z x^T)(z^T x)$

$= (z^T x)^T (z^T x)$

$= \| z^T x \|_2^2 = \sum_{i=1}^n (z_i x_i)^2 \geq 0$ for all $x \in \mathbb{R}^n$

$\therefore A = zz^T$ is positive semidefinite

COEN 240
Assignment 1

Shiva Govindaraju

2 b) $A \in \mathbb{R}^{n \times n}$ is PSD, $B \in \mathbb{R}^{m \times n}$ is arbitrary.

let $Q = BAB^T$

is Q symmetric?

$Q^T = (BAB^T)^T = B^T A^T (B^T)^T$

$= B^T A^T B = BAB^T$

since A is positive semidefinite, and thus, is symmetric so $A = A^T$.

let x be a vector of length m.

$x^T Q x = x^T (BAB^T) x = (x^T B^T) A (xB)$

$= A (xB)^T (xB)$.

$= A \cdot \|xB\|_2^2 \geq 0$ for all vectors $x \in \mathbb{R}^m$

$\therefore$ Since if A is positive semidefinite, then for $y = \|xB\|_2^2 \geq 0$, Ay is also ≥ 0.

$\therefore$ BAB^T is positive semidefinite

3 a) $J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-y^{(i)} \theta^T x^{(i)}}) = -\frac{1}{n} \sum_{i=1} \log(h_\theta(y^{(i)} x^{(i)}))$

$$\nabla J(\theta) = \frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left(\frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-y^{(i)} \theta^T x^{(i)}}) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log(1 + e^{-y^{(i)} \theta^T x^{(i)}})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{(e^{-y^{(i)} \theta^T x^{(i)}})}{(1 + e^{-y^{(i)} \theta^T x^{(i)}})} \cdot (-y^{(i)} x_j^{(i)})$$

$$= \frac{1}{n} \sum_{i=1}^{n} (1 + e^{+y^{(i)} \theta^T x^{(i)}})^{-1} (-y^{(i)} x_j^{(i)})$$

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

$$\nabla J(\theta) = \frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^{n} h_\theta(-y^{(i)} x^{(i)}) (-y^{(i)} x_j^{(i)})$$

$$\nabla^2 J(\theta) = \frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_j} J(\theta)$$

$$= \frac{\partial}{\partial \theta_k}\left(\frac{1}{n}\sum_{i=1}^{n} h_\theta(-y^{(i)}x^{(i)})(-y^{(i)}x_j^{(i)})\right)$$

$$= \frac{-1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \theta_k} h_\theta(-y^{(i)}x^{(i)})(y^{(i)}x_j^{(i)})$$

here, we note that for $g(z) = (1+e^{-z})^{-1}$

$$g'(z) = \frac{d}{dz}(1+e^{-z})^{-1} = (1+e^{-z})^{-2}(e^{-z})$$

$$= g(z)(1-g(z)).$$

we also note $h_\theta(x) = g(\theta^T x)$.

$$\therefore \frac{\partial^2 J(\theta)}{\partial\theta_k \partial\theta_j} = \frac{-1}{n}\sum_{i=1}^{n}\left(h_\theta(-y^{(i)}x^{(i)})(1-h_\theta(-y^{(i)}x^{(i)}))\right)(-y^{(i)}x_j^{(i)})(-y^{(i)}x_k^{(i)})$$

but, we may also note that $y^{(i)} \in \{-1, 1\}$ by the problem description.

$\therefore$ if $h_\theta(z)(1-h_\theta(z))$ is an even function, it can be removed to make the math a little cleaner.

$$h_\theta(-z)(1-h_\theta(-z)) = (1+e^{\theta^T z})^{-1}(1-(1+e^{\theta^T z})^{-1})$$

$$= \frac{(1+e^{\theta^T z})-1}{(1+e^{\theta^T z})^{-2}} = \frac{e^{\theta^T z}}{(1+e^{\theta^T z})^{-2}} = h_\theta(z)(1-h_\theta(z))$$

(after some more steps...)

$\therefore$ H_{jk} doesn't need $y^{(i)}$ because $y^{(i)} \in \{-1, 1\}$ and $h_\theta(z)(1-h_\theta(z))$ is an even function

$$\therefore H_{jk} = \frac{-1}{n}\sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))(x_j^{(i)}x_k^{(i)})$$

$$\boxed{H = \nabla^2 J(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))(x^{(i)}x^{(i)T})}$$

Assignment1

3 a, cont) $z^T H z = z^T \left(\frac{1}{n} \sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))(x^{(i)} x^{(i)T}) \right) z$

$= \frac{1}{n} \sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)})) \left(z^T x^{(i)} x^{(i)T} z \right)$

$= \frac{1}{n} \sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)})) (z^T x^{(i)})^2$

since $z^T x x^T z = (z^T x)(z^T x)^T = (z^T x)^2$

we know that $(z^T x^{(i)})^2 \geq 0$ for all i and z.

$h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))$

$= \frac{1}{1+e^{-\theta^T x^{(i)}}} \left(1 - \frac{1}{1+e^{-\theta^T x^{(i)}}} \right)$

as $\frac{1}{1+e^{-\theta^T x^{(i)}}} \in [0,1]$ ~~for all θ^T and x~~

we can say $h_\theta(x^{(i)})(1-h_\theta(x^{(i)})) \in [0,1]$.

$\therefore z^T H z = \frac{1}{n} \sum_{i=1}^{n} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))(z^T x^{(i)})^2 \geq 0$

$\therefore$ H is a positive semidefinite matrix.

3 b) After convergence:

$\theta = [-2.6205116, 0.76037154, 1.17194674]$

hypothesis:

$$\boxed{h_\theta(x) = g(-2.6205116 + 0.76037154 x_1 + 1.17194674 x_2)}$$

where $g(z) = (1+e^{-z})^{-1}$

c) See Attached PNG files.