Homework-1

Syeda Grousia
W11587235.

① a) Let $f(x) = \frac{1}{2} x^T A x + b^T x$; where A is a symmetric matrix and $b \in R^n$ is a vector. What is $\nabla f(x)$

sol: $\nabla f(x) = \frac{1}{2} \times 2Ax + b$ $\left[\because \frac{\partial}{\partial x} x^T A x = 2Ax \ \& \ \frac{\partial}{\partial x} b^T x = b \right]$

$\nabla f(x) = Ax + b$

b) Let $f(x) = g(h(x))$ where $g: R \to R$ is differentiable and $h: R^n \to R$ is differentiable. What is $\nabla f(x)$

$$\nabla f(x) = g'(h(x)) \frac{\partial}{\partial x} h(x)$$
$$= g'(h(x)) \nabla h(x)$$

c) Let $f(x) = \frac{1}{2} x^T A x + b^T x$, where A is symmetric and $b \in R^n$ is a vector. What is $\nabla^2 f(x)$

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

d) Let $f(x) = g(a^T x)$ where $g: R \to R$ is continuously differentiable & $a \in R^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$

$$\nabla f(x) = g'(a^T x) \cdot \frac{\partial}{\partial x} (a^T x)$$
$$= g'(a^T x) \cdot a$$
$$\nabla^2 f(x) = g''(a^T x) \cdot a \cdot \frac{\partial}{\partial x} a^T x$$
$$= g''(a^T x)\, a \cdot a^T$$

(2) a) Let $Z \in R^n$ be an n length vector. Show that $A = ZZ^T$ is positive definite.

$$A = ZZ^T$$

A matrix A is positive definite if $x^T A x > 0$ $\forall x \neq 0$ and $A^T = A$.

So $x^T A x = x^T Z Z^T x$

$= (x^T Z)(Z^T x)$ $[\because Z^T = Z \text{ \& } x = x^T]$

$= (x^T Z)^2 \geq 0$

$\because A = ZZ^T$ is positive semidefinite.

b) BAB^T

$x^T A x > 0$

substituting BAB^T in A we get

$\cancel{P}\, x^T B A B^T x$

$= (x^T B) A (B^T x)$

$= (x B^T)^T A (B^T x)$ $[\because B \in R^{m \times n}]$

$= z^T A z \geq 0$ $[z = B^T x \in R^n$ & since A is PSD$]$

$\therefore$ BAB^T is a positive semidefinite matrix.

③ The Avg empirical loss for logistic regression is given as

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

where $y^{(i)} \in \{-1, 1\}$, $h_\theta(x) = g(\theta^T x)$ & $g(z) = \frac{1}{1+e^{-z}}$

show that $Z^T H Z \geq 0$

Sol: Hessian H is the second derivative of the function.

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

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

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

$$= -\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+e^{y^{(i)}\theta^T x^{(i)}}} \left(-y^{(i)} x^{(i)}\right) \qquad \left[\because \frac{e^{-y^{(i)}\theta^T x^{(i)}}}{1+e^{-y^{(i)}\theta^T x^{(i)}}} = \frac{1}{1+e^{y^{(i)}\theta^T x^{(i)}}}\right]$$

$$= +\frac{1}{n}\sum_{i=1}^{n} g\left(-y^{(i)}\theta^T x^{(i)}\right)\left(y^{(i)} x^{(i)}\right)$$

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

$$\nabla^2 J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\left(h_\theta\left(-y^{(i)} x^{(i)}\right)\right)\left(y^{(i)} x^{(i)}\right)$$

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

$$= -\frac{1}{n}\sum_{i=1}^{n} \left(y^{(i)} x^{(i)}\right)^2 \left(h_\theta\left(-y^{(i)} x^{(i)}\right)\right)\left(1-h_\theta\left(-y^{(i)} x^{(i)}\right)\right)$$

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

$$\therefore H = \frac{1}{n}\sum_{i=1}^{n} h(x^{(i)})(1-h(x^{(i)}))\, x^{(i)} x^{(i)T} \quad \left[\because y^{(i)} \in \{-1,1\} \;\&\; h(-x) = h(x)\right]$$

For any vector z, it holds that $z^T H z \geq 0$ that is to show that H is positive semidefinite

$$z^T H z$$

$$= z^T \cdot \left[\frac{1}{n}\sum_{i=1}^{n} h(x^{(i)})(1-h(x^{(i)}))\, x^{(i)} x^{(i)T}\right] \cdot z$$

$$= \frac{1}{n}\sum_{i=1}^{n} h(x^{(i)})(1-h(x^{(i)}))\, z^T x^{(i)} x^{(i)T} z$$

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

Since $h(x^{(i)})(1-h(x^{(i)}))$ is positive & $(z^T x^{(i)})^2 \geq 0$ we can say that H is positive semidefinite.

3 b) The coefficients θ that were obtained after the fit are

$\theta[0] = -2.62051159718 01973$
$\theta[1] = 0.7603715358976769$
$\theta[2] = 1.1719467415671432$