HOMEWORK - 1

CS 541 DEEP LEARNING

Q 3.

a) Prove

$$\nabla_x (x^T a) = \nabla_x (a^T x) = a$$

x, a & column vectors.

$\nabla_x (x^T a)$.

Let x be $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1}$ & a be $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{n \times 1}$

$$x^T = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}_{1 \times n}$$

$$x^T a = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}_{1 \times n} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{n \times 1}$$

$$x^T a = \begin{bmatrix} x_1 a_1 + x_2 a_2 + \cdots + x_n a_n \end{bmatrix}_{1 \times 1}$$

$$\nabla_x = \begin{bmatrix} \frac{\partial f}{x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{x_n} \end{bmatrix}_{n\times 1} \begin{bmatrix} \cancel{x_1 a_1} \end{bmatrix} \begin{bmatrix} x_1 a_1 + x_2 a_2 + \dots + x_n a_n \end{bmatrix}_{1\times 1}$$

here $f = x^T a$.

$$\therefore \nabla_x = \begin{bmatrix} \frac{\partial (x_1 a_1 + x_2 a_2 + x_3 a_3 + \dots x_n a_n)}{\partial x_1} \\ \frac{\partial (x_1 a_1 + x_2 a_2 + \dots x_n a_n)}{\partial x_2} \\ \vdots \\ \frac{\partial (x_1 a_1 + x_2 a_2 + \dots + x_n a_n)}{\partial x_n} \end{bmatrix}_{n\times 1}$$

$$\therefore \nabla_x = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}_{n\times 1} \rightarrow (1)$$

$$= a$$

?

$$a^T x = [a_1 \; a_2 \; \cdots \; a_n]_{1 \times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$a^T x = [a_1 x_1 + a_2 x_2 + \cdots \; a_n x_n]^{n \times 1}$$

$$\nabla x = \begin{bmatrix} \dfrac{\partial (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)}{\partial x_1} \\ \partial (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) \\ \vdots \\ \dfrac{\partial (a_n x_n + a_{n-1} x_{n-1} + \cdots + a_2 x_2 + a_1 x_1)}{\partial x_n} \end{bmatrix}$$

$$\therefore \nabla x = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = a \rightarrow (2)$$

From (1) & (2) we can infer

$$\boxed{\nabla x (x^T a) = \nabla x (a^T x) = a}$$

where x & a are column vectors

b) Prove that

$$\nabla_x (x^T A x) = (A + A^T) x$$

Let x be $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}_{n \times 1}$ a column vector (given)

and $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{1n} & \cdots & \cdots & a_{nn} \end{bmatrix}_{n \times n}$

$$x^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}_{1 \times n} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}_{n \times n}$$

$\cancel{x^T A = \Big[}$

$$x^T A = \left[x_1 a_{11} + x_2 a_{21} + \cdots + x_n a_{n1} \quad \cdots\cdots \quad x_1 a_{1n} + x_2 a_{2n} + \cdots + x_n a_{nn} \right]_{1\times n}$$

$$x^T A x = \left[x_1 a_{11} + \cdots + x_n a_{n1} \quad \cdots\cdots \quad x_1 a_{1n} + \cdots + a_n a_{nn} \right]_{1\times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$$= \left[x_1 (x_1 a_{11} + x_2 a_{21} \cdots + a_{n1} x_n) + x_2 (x_1 a_{12} + \cdots + x_n a_{n2}) + \cdots \cdots x_n (x_1 a_{1n} + \cdots + x_n a_{nn}) \right]_{1\times 1}$$

$$\nabla_x (x^T A x) = \begin{bmatrix} \dfrac{\partial ([x_1(x_1 a_{11} + \cdots + x_n a_{n1}) + \cdots\cdots + x_n(x_1 a_{1n} + \cdots + x_n a_{nn})])}{\partial x_1} \\ \vdots \\ \dfrac{\partial ([x_1(x_1 a_{11} + \cdots + x_n a_{n1}) + \cdots\cdots + x_n(x_1 a_{1n} + \cdots + x_n a_{nn})])}{\partial x_n} \end{bmatrix}$$

$$2a_{11}x_1 + a_{21}x_2 + \cdots + x_n a_{n1}$$

$$a_{12}x_1 + 2a_{22}x_2 + \cdots + x_n a_{n2}$$

$$\vdots$$

$$= \cancel{(2a_{11}x_1 + a_{21}x_2 + \cdots x_n a_{n1}) + x_2 a_{12} + x_3 a_{13} + \cdots + x}$$

$$= \begin{bmatrix} (2a_{11}x_1 + a_{21}x_2 + \cdots + x_n a_{n1}) + x_2 a_{12} + \cdots + x_n a_{1n} \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 \cdots + (x_1 a_{1n} + \cdots + 2a_{nn}x_n) \end{bmatrix}_{n\times 1}$$

$$= \begin{bmatrix} 2a_{11}x_1 + x_2(a_{12}+a_{21}) + \cdots + x_n(a_{1n}+a_{n1}) \\ \vdots \\ x_1(a_{1n}+a_{n1}) + x_2(a_{2n}+a_{n2}) + \cdots + 2a_{nn}x_n \end{bmatrix}_{n\times 1} \rightarrow (1)$$

$A + A^T$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix}_{n\times n} + \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ \vdots & & \ddots & \vdots \\ a_{1n} & \cdots & \cdots & a_{nn} \end{bmatrix}_{n\times n}$$

$$= \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & 2a_{22} & \cdots & a_{2n}+a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{n1}+a_{1n} & \cdots & \cdots & 2a_{nn} \end{bmatrix}_{n\times n}$$

$(A + A^T)x$

$$= \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & \cdots & \cdots & a_{2n}+a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{n1}+a_{n1} & \cdots & \cdots & 2a_{nn} \end{bmatrix}_{n\times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$$\therefore (A+A^T)x$$

$$= \begin{bmatrix} 2a_{11}x_1 + (a_{12}+a_{21})x_2 - - - - + (a_{1n}+a_{n1})x_n \\ \vdots \\ (a_{1n}+a_{n1})x_n + (a_{2n}+a_{n2})x_2 + - - - - - - 2a_{nn}x_n \end{bmatrix}_{n\times 1} \rightarrow (2)$$

From (1) & (2)

we can infer that (1) & (2) are equal

$$\boxed{\therefore \nabla_x (x^T A x) = (A+A^T)x}$$

c) Prove that $\nabla_x (x^T A x) = 2Ax$ if A is a symmetric matrix.

From the proof from question (b)

we can say that

$$\nabla_x (x^T A x) = (A+A^T)x \rightarrow (1)$$

For a symmetric matrix $A^T = A$.

$\therefore A^T + A = A + A = 2A$.

$\therefore (A^T + A)x = 2Ax$

From ① we can say that

$$\boxed{\nabla_x (x^T A x) = 2Ax}$$

d) Based on the theorems above, prove that

$\nabla_x [(Ax+b)^T (Ax+b)] = 2A^T(Ax+b)$
where A is $n \times n$ symmetric matrix

LHS:

$(Ax+b)^T = (Ax)^T + b^T$
$= x^T A^T + b^T$

$(Ax+b)^T (Ax+b) = (x^T A^T + b^T)(Ax+b)$
$= x^T A^T A x + x^T A^T b$
$+ \cancel{x} b^T A x + b^T b$
$= x^T A A x + x^T A b$
$+ (A^T b)^T x + b^T b$

$\therefore A^T = A$ since A is symmetric.

$$= x^T AAx + x^T Ab + (Ab)^T x + b^T b.$$

$\therefore A^T = A$ ~~ia~~ since A is symmetric

$$\nabla\left[(Ax+b)^T (Ax+b)\right]$$

$$= \nabla(x^T AAx) + \nabla(x^T Ab) + \nabla((Ab)^T x) + \cancel{b^T b}\,\nabla(b^T b)$$

WRT $\nabla(x^T A x) = (A + A^T)x$

$\nabla(x^T a) = a = \nabla(a^T x)$

& $\nabla(\text{constant}) = 0$

$$\therefore \nabla((Ax+b)^T (Ax+b))$$

$$= (AA + (AA)^T)x + Ab + Ab + 0.$$

$$= (AA + AA)x + Ab + Ab$$

$$= AAx + AAx + Ab + Ab$$

$$= \cancel{2}A(Ax + Ax + b + b)$$

$$= 2A(Ax + b)$$

$$\nabla_x\left[\cancel{(Ax+b)^T (Ax+b)}\right] = \boxed{2A^T(Ax+b) = \nabla_x\left[(Ax+b)^T (Ax+b)\right]}$$

$\therefore A^T = A$, A is a symmetrical matrix