Sumeet Shanbhag

Deep Learning Proofs and Results

`

```
PS C:\Users\Sumeet\OneDrive - Worcester Polytechnic Institute (wpi.edu)\WPI Sem 2\Dee
p Learning\Homework> & C:/Users/Sumeet/AppData/Local/Programs/Python/Python311/python
.exe "c:/Users/Sumeet/OneDrive - Worcester Polytechnic Institute (wpi.edu)/WPI Sem 2/
Deep Learning/Homework/homework1_shanbhag_code.py"
Mean Squared Error for the testing set:  206.79647491430478
Mean Squared Error for the training set:  39.242962989290696
PS C:\Users\Sumeet\OneDrive - Worcester Polytechnic Institute (wpi.edu)\WPI Sem 2\Dee
p Learning\Homework> 
```

As we can see, we get the following values for testing and training set

Mean Squared Error for the testing set: 206.79647491430478

Mean Squared Error for the training set: 39.242962989290696

Sumeet Shanbhag

Deep Learning Proofs.

3 a. Prove that $\nabla_x (x^T.a) = \nabla_x (a^T x) = a$.

~~$x^T a =$~~

$$\nabla_x \text{ is given by } \nabla_x = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

And $x^T a$ is given by $\Rightarrow [x_1 a_1 + x_2 a_2 + .. x_n a_n] = F$

$$F = x^T.a \text{ which means } \nabla_x \to \begin{bmatrix} \frac{\partial [x_1 a_1 + \ldots + x_n a_n]}{\partial x_1} \\ \vdots \\ \frac{\partial [x_1 a_1 + \ldots + x_n a_n]}{\partial x_n} \end{bmatrix}$$

After derivation,

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a.$$

Doing the same step on RHS,

$$F = a^T x = [a_1 x_1 + a_2 x_2 .. + a_n x_n]$$

$$\text{So } \nabla_x \Rightarrow \begin{bmatrix} \frac{\partial [a_1 x_1 + .. + a_n x_n]}{\partial x_1} \\ \vdots \\ \frac{\partial [a_1 x_1 + \ldots + a_n x_n]}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

This shows that LHS = RHS = a.

3.b) Prove that $\nabla_x (x^T A x) = (A + A^T)x$

Since A is a $n \times n$ matrix & $x \in \mathbb{R}^n$

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

$$\cancel{= [(x_1 a_{11} + x_2 a_{21} + \ldots} \; x_2 a_{21} + \ldots + (a x_n a_{1n}$$

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

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

$$F = \left[x_1 (x_1 a_{11} .. + x_n a_{n1}) + .. + x_n (x_1 a_{1n} .. + x_n a_{nn}) \right]_{1 \times 1}.$$

Now $\nabla_x (x^T A x) = \begin{bmatrix} \partial F / \partial x_1 \\ \vdots \\ \partial F / \partial x_n \end{bmatrix}$

$$= \begin{bmatrix} \dfrac{\partial [x_1 (x_1 a_{11} + .. + x_n a_{n1}) + .. + x_n (x_1 a_{1n} + .. + x_n a_{nn})]}{\partial x_1} \\ \vdots \\ \partial F / \partial x_n \end{bmatrix}$$

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

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

$RHS = A + A^T$

$$= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ \vdots & & & \\ a_{1n} & \dots & & a_{nn} \end{bmatrix}$$

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

$$(A+A^T)x = \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \dots & a_{1n}+a_{n1} \\ \vdots & & & \\ a_{n1}+a_{1n} & & & 2a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} 2a_{11}x_1 + x_2(a_{12}+a_{21}) + x_n(a_{1n}+a_{n1}) \\ \\ x_1(a_{1n}+a_{n}) \dots + 2a_{nn}x_n \end{bmatrix}$$

We can see that $LHS = RHS$.

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

3.c) Prove that

$$\nabla_x (x^T A x) = 2Ax.$$

We know that for a symmetric matrix,

$$A = A^T$$

which mean, $A + A^T = A + A = 2A$ (i)

In the previous derivation, we proved that

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

From (i), we can substitute

$$x^T A x = 2Ax.$$

Hence proved.

3.d) Prove that

$$\nabla_x [(Ax + b)^T . (Ax + b)] = 2A^T (Ax + b)$$

LHS

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

$$\Rightarrow x^T A^T + b^T$$

$$(Ax + b)^T . (Ax + B) \Rightarrow (x^T A^T + b^T) . (Ax + b)$$

$$\Rightarrow x^T A^T A x + x^T A^T b + b^T A x + b^T b.$$

We know that $A^T = A$

The expression becomes,

$$x^T A A x + x^T A b + b^T A x + b^T b$$

$$x^T A A x + x^T A b + (Ab)^T x + b^T b.$$

$$\nabla_x [(Ax+b)^T (Ax+b)] \Rightarrow \nabla_x [x^T A A x + x^T A b + (Ab)^T x + b^T b]$$

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

From the derived proofs, we can rewrite the expression as.

$$= [AA + (AA)^T] x + Ab + Ab + 0$$

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

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

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

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

$$= RHS$$

$\therefore$ LHS = RHS = Hence Proved.