Vaibhav Nandkumar Kadam
vkadam@wpi.edu

Aadesh Varude
avarude@wpi.edu

Q2 . Solution.

Note : The function `linear_regression(X_tr, y_tr)` returns weight vector the last element of the weight vector is the bias.
The cost fMSE on the training set Dtr is 39.24296299.
The cost fMSE on the training set Dte is 206.79647485.

Q3 a. Solution.

We have x and a are any two column vectors, we have,

$$x^T = \begin{bmatrix} x_1 & x_2 & \cdot\cdot & x_n \end{bmatrix} \tag{1}$$

$$a^T = \begin{bmatrix} a_1 & a_2 & \cdot\cdot & a_n \end{bmatrix} \tag{2}$$

$$x^T a = \begin{bmatrix} x_1 & x_2 & \cdot\cdot & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = x_1 a_1 + x_2 a_2 + x_3 a_3 \cdots x_n a_n \tag{3}$$

Similarly,

$$a^T x = a_1 x_1 + a_2 x_2 + a_3 x_3 \cdots a_n x_n \tag{4}$$

$$\nabla_x (x^T a) = \begin{bmatrix} \frac{\partial x^T a)}{x_1} \\ \cdot \\ \cdot \\ \frac{\partial x^T a)}{x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = a \tag{5}$$

Similarly,

$$\nabla_x (a^T x) = a \tag{6}$$

Q3 b. Solution.

We have,

$$x^T A x = \begin{bmatrix} x_1 & x_2 & \cdot\cdot & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdot\cdot & a_{1n} \\ \cdot & & \\ \cdot & & \\ a_{n1} & \cdot\cdot & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{7}$$

$$x^T A x = \begin{bmatrix} (a_{11} x_1 \cdots a_{n1} x_n) & \cdot & \cdot & (a_{1n} x_1 \cdots a_{nn} x_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{8}$$

$$x^T Ax = \begin{bmatrix} \Sigma_{i=1}^n a_{i1} x_i & \cdot & \cdot & \Sigma_{i=1}^n a_{in} x_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{9}$$

$$x^T Ax = \begin{bmatrix} x_1 \Sigma_{i=1}^n a_{i1} x_i & \cdot & \cdot & x_n \Sigma_{i=1}^n a_{in} x_i \end{bmatrix} \tag{10}$$

$$x^T Ax = \Sigma_{j=1}^n x_j \Sigma_{i=1}^n a_{ij} x_i \tag{11}$$

$$x^T Ax = \Sigma_{j=1}^n \Sigma_{i=1}^n a_{ij} x_i x_j \tag{12}$$

$$\frac{\partial x^T Ax}{x} = \begin{bmatrix} \frac{\partial}{x_1} (\Sigma_{i=1}^n a_{ij} x_i x_j) \\ \frac{\partial}{x_2} (\Sigma_{i=1}^n a_{ij} x_i x_j) \\ \cdot \\ \cdot \\ \frac{\partial}{x_n} (\Sigma_{i=1}^n a_{ij} x_i x_j) \end{bmatrix} \tag{13}$$

$$\frac{\partial x^T Ax}{x} = \begin{bmatrix} \Sigma_{j=1}^n a_{1j} x_j) + \Sigma_{i=1}^n a_{i1} x_i) \\ \Sigma_{j=1}^n a_{2j} x_j) + \Sigma_{i=1}^n a_{i2} x_i) \\ \cdot \\ \cdot \\ \Sigma_{j=1}^n a_{nj} x_j) + \Sigma_{i=1}^n a_{in} x_i) \end{bmatrix} = \begin{bmatrix} \Sigma_{j=1}^n a_{1j} x_j) \\ \Sigma_{j=1}^n a_{2j} x_j) \\ \cdot \\ \cdot \\ \Sigma_{j=1}^n a_{nj} x_j) \end{bmatrix} \begin{bmatrix} \Sigma_{i=1}^n a_{i1} x_i) \\ \Sigma_{i=1}^n a_{i2} x_i) \\ \cdot \\ \cdot \\ \Sigma_{i=1}^n a_{in} x_i) \end{bmatrix} \tag{14}$$

$$\frac{\partial x^T Ax}{x} = Ax + A^T x \tag{15}$$

Thus,

$$\boldsymbol{\nabla}_x (x^T Ax) = Ax + A^T x \tag{16}$$

Hence Proved.

Q3 c. Solution.

We have from Q3.b eq. 16

$$\boldsymbol{\nabla}_x (x^T Ax) = Ax + A^T x \tag{17}$$

For a symmetric matrix A we know

$$A_T = A \tag{18}$$

$$\boldsymbol{\nabla}_x (x^T Ax) = (A + A)x = 2Ax \tag{19}$$

Hence Proved.

Q3 d. Solution.

$$\boldsymbol{\nabla}_x [(Ax + b)^T (Ax + b)] \tag{20}$$

$$(Ax + b)^T = (Ax)^T + b^T = x^T A^T + b^T \tag{21}$$

$$(Ax + b)^T (Ax + b) = (x^T A^T + b^T)(Ax + b) \tag{22}$$

$$= (x^T A^T Ax + x^T A^T b + b^T Ax + b^T b) \tag{23}$$

$$\boldsymbol{\nabla}_x [(Ax + b)^T (Ax + b)] = \boldsymbol{\nabla}_x (x^T A^T Ax) + \boldsymbol{\nabla}_x (x^T A^T b) + \boldsymbol{\nabla}_x (b^T Ax) + \boldsymbol{\nabla}_x (b^T b) \tag{24}$$

We get from Q 3.b,

$$\boldsymbol{\nabla}_x(x^T(A^TA)x) = (A^TA + AA^T)x \tag{25}$$

As A is symmetric,

$$= (A^TA + A^TA)x \tag{26}$$

Thus,

$$\boldsymbol{\nabla}_x[(Ax+b)^T(Ax+b)] = 2A^TAx \tag{27}$$

From Q3. a

$$\boldsymbol{\nabla}_x(x^T(A^Tb)) = A^Tb \tag{28}$$

As A is symmetric,

$$b_T(Ax) = (Ax)^Tb \tag{29}$$

$$\boldsymbol{\nabla}_x(b^TAx) = \boldsymbol{\nabla}\big(x^TA^Tb\big) \tag{30}$$

$$\boldsymbol{\nabla}_x(b^TAx) = \boldsymbol{\nabla}\big(x^TA^Tb\big) \tag{31}$$

We have,

$$\boldsymbol{\nabla}_x(b^Tb) = 0 \tag{32}$$

Substituting above equations in eq. 24

$$\boldsymbol{\nabla}_x[(Ax+b)^T(Ax+b)] = 2A^TAx + 2A^Tb = 2A^T(Ax+b) \tag{33}$$