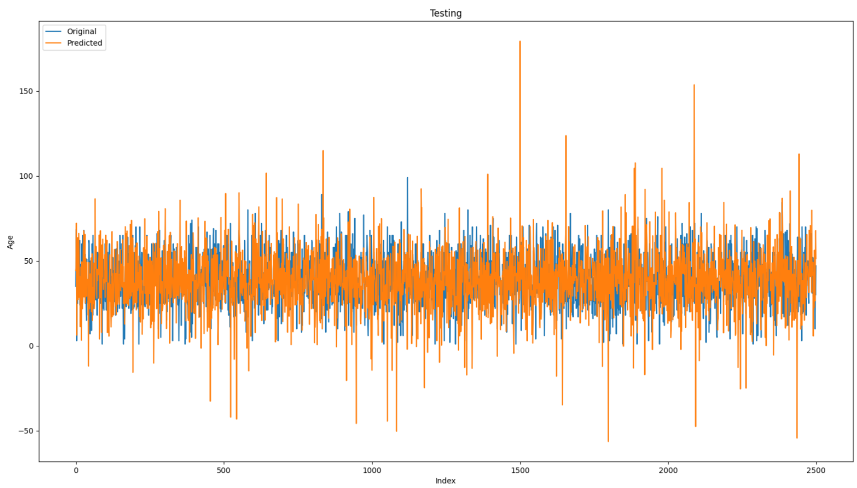
2a).



Actual vs prediction for training and testing data

RMSE:

Training Error = 39.243

Testing Error = 206.796

3 a) $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $\quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

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

$$x^T a = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \quad — (1)$$

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

$$a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \quad — (2)$$

From (1) & (2):

$$x^T a = a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

$$\nabla_x (x^T a) = \nabla_x (a^T x) = \begin{bmatrix} \frac{\partial}{\partial x_1} [a_1 x_1 + a_2 x_2 + \cdots + a_n x_n] \\ \frac{\partial}{\partial x_2} [a_1 x_1 + a_2 x_2 + \cdots + a_n x_n] \\ \vdots \\ \frac{\partial}{\partial x_n} [a_1 x_1 + a_2 x_2 + \cdots + a_n x_n] \end{bmatrix}$$

$$\nabla_x (x^T a) = \nabla_x (a^T x) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

3.b) Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$x^T A x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

$$x^T A x = x_1(a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n) + x_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n)$$
$$+ \cdots + x_n(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n)$$

$$\nabla_x (x^T A x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x^T A x) \\ \frac{\partial f}{\partial x_2}(x^T A x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x^T A x) \end{bmatrix}$$

UV

$UV' + UV'$

$x_1(a_{11})$

$$\nabla_x (x^T A x) = \begin{bmatrix} (a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n) + a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n \\ a_{12}x_1 + (a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n) + a_{22}x_2 + \cdots + a_{n2}x_n \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + (a_{n1}x_1 + a_{n2}x_2 + \cdots a_{nn}x_n) + a_{nn}x_n \end{bmatrix}$$

$$\nabla_x (x^T A x) = \begin{bmatrix} 2a_{11}x_1 + (a_{12}+a_{21})x_2 + \cdots + (a_{1n}+a_{2n})x_n \\ (a_{21}+a_{12})x_1 + 2a_{22}x_2 + \cdots + (a_{2n}+a_{n2})x_n \\ \vdots \\ (a_{n1}+a_{1n})x_1 + (a_{n2}+a_{2n})x_2 + \cdots + 2a_{nn}x_n \end{bmatrix} \quad (3)$$

$$(A + A^T)x = \left\{ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{2n} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \right\} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{2n} \\ a_{21}+a_{12} & 2a_{22} & \cdots & a_{2n}+a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{n1}+a_{1n} & a_{n2}+a_{2n} & & 2a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$(A+A^T)x = \begin{bmatrix} 2a_{11}x_1 + (a_{12}+a_{21})x_2 + \cdots + (a_{1n}+a_{2n})x_n \\ (a_{21}+a_{12})x_1 + 2a_{22}x_2 + \cdots + (a_{2n}+a_{n2})x_n \\ \vdots \\ (a_{n1}+a_{1n})x_1 + (a_{n2}+a_{2n})x_2 + \cdots + 2a_{nn}x_n \end{bmatrix} \quad \text{--- (4)}$$

From (3) and (4)

$$\boxed{\nabla_x (x^T A x) = (A+A^T)x} \quad \text{--- (5)}$$

3.c) If a matrix A is symmetric, then the following condition is satisfied

$$A = A^T$$

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

From (5) $\rightarrow$ $(A+A^T)x = 2Ax$

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

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

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

$(Ax+b)^T(Ax+b) = x^TA^TAx + x^TA^Tb + b^TAx + b^Tb$

x^TA^Tb is same as b^TAx

$(Ax+b)^T(Ax+b) = x^TA^TAx + 2x^TA^Tb + b^Tb$

Let $A^TA = B$ and $A^Tb = C$

$(Ax+b)^T(Ax+b) = x^TBx + 2x^TC + b^Tb$

$\nabla_x[(Ax+b)^T(Ax+b)] = \nabla_x(x^TBx) + \nabla_x(2x^TC) + \nabla_x(b^Tb)$

From (b) : $\nabla_x(x^TBx) = (B+B^T)x$

From 3.a : $\nabla_x(2x^TC) = 2\nabla_x(x^TC) = 2C$

Since b is a constant column vector $\nabla_x(b^Tb) = 0$

$\therefore \nabla_x[(Ax+b)^T(Ax+b)] = (B+B^T)x + 2C$

$= (A^TA + AA^T)x + 2A^Tb$

And it is given that A is symmetric

That means A^TA is same as AA^T

$$\nabla_x \left[(Ax+b)^T (Ax+b) \right] = 2A^T A x + 2A^T b$$

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

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