$X_{dim} = (3,2)$ | $w_{dim} = ($ $)$

1.1

House

SQFT : 2,000 x_1

Bedrooms : 3 x_2

weights

$[150 \;\; 10000]$ (\$/sqft, \$/bedroom)

bias

$b = 50{,}000$

A. Dot Product

$$w \bullet x = 150 \cdot 2000 + 3 \cdot 10000$$
$$= 30000 + 30000$$
$$\Rightarrow \hat{y} = w \bullet x + b = 60000 + 50000$$
$$\Rightarrow \hat{y} = 110000$$

A 2,000 sqft house with three bedrooms would cost \$110,000.

B. Multiple Predictions

$n = 3$ houses

$$X = \begin{array}{cc} \text{sqft} & \text{beds} \\ \begin{bmatrix} 2000 & 3 \\ 1500 & 2 \\ 2500 & 4 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

Keep w & b the same.

$$X \cdot w^T = \begin{bmatrix} 2000 & 3 \\ 1500 & 2 \\ 2500 & 4 \end{bmatrix} \begin{bmatrix} 150 \\ 10000 \end{bmatrix}$$

$$= \begin{bmatrix} 2000(150) + 3(10000) \\ 1500(150) + 2(10000) \\ 2500(150) + 4(10000) \end{bmatrix} = \begin{bmatrix} 30000 + 30000 \\ 22500 + 20000 \\ 37500 + 40000 \end{bmatrix} = \begin{bmatrix} 60000 \\ 42500 \\ 77500 \end{bmatrix}$$

$$\Rightarrow \hat{y} = Xw^T + b = \begin{bmatrix} 60000 \\ 42500 \\ 77500 \end{bmatrix} + \begin{bmatrix} 50000 \\ 50000 \\ 50000 \end{bmatrix} = \begin{bmatrix} 110000 \\ 92500 \\ 127500 \end{bmatrix}$$

1.2 For $A = (m \times n)$, $B = (p \times q)$,

A.

AB iff $n = p$ (row x col) $\Rightarrow$ col = col

- $(100 \times 5)(5 \times 1) \rightarrow$ Valid, $(100, 1)$
- $(100 \times 5)(3 \times 1) \rightarrow$ Invalid
- $(1 \times 5)(5 \times 100) \rightarrow$ Valid, (1×100)
- $(100 \times 1)(1 \times 5) \rightarrow$ Valid, (100×5)

B.

$I = 5$ (observations)

$K = 2$ (features)

out $= 1$ (MPG)

$X : (5 \times 2)$

$y : (5 \times 1)$

$e : (5 \times 1)$

$w : (1 \times 2)$

$$L = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_i (e)^2$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial e} \cdot \frac{\partial e}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

$$\Rightarrow \frac{\partial}{\partial e} L = \frac{1}{n} \frac{\partial}{\partial e} (e)^2 = \frac{1}{n} \cdot 2e = \frac{2}{n} e$$

$$\Rightarrow \frac{\partial}{\partial \hat{y}} e = \frac{\partial}{\partial \hat{y}} (y - \hat{y}) = -1$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial w} = \frac{\partial}{\partial w} (wx + b) = x$$

Since we're working with vectors, $e \to e^T$

$$\frac{2}{n} e^T \bullet -1 \bullet X$$

$$= -\frac{2}{n} e^T X$$

Q.E.D.

$$\Rightarrow e^T_{dim} = (1, 5)$$
$$\Rightarrow (e^T X)_{dim} = (1, 2)$$
$$\Rightarrow (e^T X)_{dim} = (w)_{dim}$$

This operation accomplishes gradient descent because each operation is based on the same values as the loop approach. When we multiply the matrix X by e^T, it is the same as $x_1 \cdot e_1 + x_2 e_2 + \cdots + x_i e_i$ since dot products take the same shape. The (1×2) result is the relative effect of each weight on the loss function in vector form. By adding them together, we see the collective effect that "votes" for the direction to change

---

Part 2

$$X = \begin{bmatrix} 3.0 & 1.1 \\ 2.5 & 0.9 \\ 4.0 & 1.5 \\ 3.5 & 1.2 \\ 2.8 & 1.0 \end{bmatrix} \quad y = \begin{bmatrix} 25 \\ 30 \\ 18 \\ 22 \\ 28 \end{bmatrix}$$

weight, hp

x_1 , x_2

$$w^T : \begin{bmatrix} 1.2 \\ -0.8 \end{bmatrix} \quad \begin{array}{l} b = 17 \\ \alpha = 0.01 \\ n = 5 \end{array}$$

$$\hat{y} = w^T X + b$$

$$\Rightarrow w^T X = \begin{bmatrix} 3.0 & 1.1 \\ 2.5 & 0.9 \\ 4.0 & 1.5 \\ 3.5 & 1.2 \\ 2.8 & 1.0 \end{bmatrix} \begin{bmatrix} 1.0 \\ -0.8 \end{bmatrix} = \begin{bmatrix} (3.0 \cdot 1.0) + (1.1 \cdot -0.8) \\ (2.5 \cdot 1.0) + (0.9 \cdot -0.8) \\ (4.0 \cdot 1.0) + (1.5 \cdot -0.8) \\ (3.5 \cdot 1.0) + (1.2 \cdot -0.8) \\ (2.8 \cdot 1.0) + (1.0 \cdot -0.8) \end{bmatrix} = \begin{bmatrix} 2.12 \\ 1.78 \\ 2.8 \\ 2.54 \\ 2 \end{bmatrix}$$

$$\Rightarrow \hat{y} = \begin{bmatrix} 2.12 \\ 1.78 \\ 2.80 \\ 2.54 \\ 2.00 \end{bmatrix} + [17] = \begin{bmatrix} 19.12 \\ 18.78 \\ 19.80 \\ 19.54 \\ 19.00 \end{bmatrix}$$ ← Predicted MPG

$$\Rightarrow \frac{\partial L}{\partial w} = -\frac{2}{n} e^T X$$

$$e = y - \hat{y}$$

$$\Rightarrow e^T = \begin{bmatrix} 25 - 19.12 \\ 30 - 18.78 \\ 18 - 19.80 \\ 22 - 19.54 \\ 25 - 19.00 \end{bmatrix} = \begin{bmatrix} 5.88 \\ 11.22 \\ -1.8 \\ 2.46 \\ 9 \end{bmatrix}$$

$(5, 1)$

$$\Rightarrow e^T X = \begin{bmatrix} 3.0 \cdot 5.88 & 1.1 \cdot 5.88 \\ 2.5 \cdot 11.22 & 0.9 \cdot 11.22 \\ 4.0 \cdot -1.8 & 1.5 \cdot -1.8 \\ 3.5 \cdot 2.46 & 1.2 \cdot 2.46 \\ 2.8 \cdot 9 & 1.0 \cdot 9 \end{bmatrix}$$

$\overline{72.3}$ $\overline{25.816}$

$x_{dim} = (2, 5)$

$e^T X_{dim} = (2, 1)$

Gradient Descent

$$W_{n+1} = W_n - \alpha \frac{\partial L}{\partial w}$$

$$\Rightarrow W_{n+1} = \begin{bmatrix} 1.2 \\ -0.8 \end{bmatrix} - \left(0.01 \cdot \begin{bmatrix} 72.3 \\ 25.818 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1.2 \\ -0.8 \end{bmatrix} - \left(\begin{bmatrix} 0.723 \\ 0.25818 \end{bmatrix} \right)$$

$$\Rightarrow W_{n+1} = \begin{bmatrix} 0.477 \\ -1.05818 \end{bmatrix}$$

Reduce both w_1 & w_2.