Lecture 1.7

- Linear Regression using Matrices
- Multiple Linear Regression

Linear Regression using Matrices

- We aim to model the relationship between input variables $\boldsymbol{X}$ and output y using a linear model:

$$y = \boldsymbol{XW} + \varepsilon$$

Where,

y:Dependent/Output variable

$\boldsymbol{X}$:Independent variable/feature set

$\boldsymbol{W}$:Coefficients/Unknown variables

ε:Noise

Cost Function Minimization

- Cost Function: $E(W) = \frac{1}{2m}\|y - XW\|^2$
- Cost Function Expansion: $E(W) = \frac{1}{2m}(y - XW)^T(y - XW)$
- Simplify: $E(W) = \frac{1}{2m}(y - XW)^T(y - XW)$

$E(W) = \frac{1}{2m}(yy^T - 2W^TX^Ty + X^TW^TXW)$

- Gradient: $\frac{\delta E(W)}{\delta W} = \frac{1}{2m}(-2X^Ty + 2X^TXW)$
- Set the gradient to zero to minimize:

$$-X^Ty + X^TXW = 0$$

Finally,

$$\boldsymbol{W} = \left(\boldsymbol{X^TX}\right)^{-\mathbf{1}}\boldsymbol{X^Ty}$$

Example 1

Consider X_0 for w_0	X_1	Y
1	1	2
1	2	3
1	3	5

- $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, Y = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, Find $y = w_0 + w_1 x_1$

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{w_0} \\ \boldsymbol{w_1} \end{bmatrix} = (\boldsymbol{X^T X})^{-\mathbf{1}} \boldsymbol{X^T y} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \right)^{-\mathbf{1}} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} & \mathbf{3} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

Solution

- $\boldsymbol{X^T X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1.1 + 1.1 + 1.1 & 1.1 + 1.2 + 1.3 \\ 1.1 + 2.1 + 3.1 & 1.1 + 2.2 + 3.3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$
- $(X^T X)^{-1} = \frac{1}{3.14 - 6.6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} 0.875 & -0.375 \\ -0.375 & 0.1875 \end{bmatrix}$
- $X^T y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 1.2 + 1.3 + 1.5 \\ 1.2 + 2.3 + 3.5 \end{bmatrix} = \begin{bmatrix} 10 \\ 23 \end{bmatrix}$
- $W = \begin{bmatrix} 0.875 & -0.375 \\ -0.375 & 0.1875 \end{bmatrix} \begin{bmatrix} 10 \\ 23 \end{bmatrix} = \begin{bmatrix} 0.875.10 & + & -0.375.23 \\ -0.375.10 & & + 0.1875.23 \end{bmatrix} = \begin{bmatrix} 0.1250 \\ 0.5625 \end{bmatrix}$

Multiple Linear Regression

- **Multiple Linear Regression** is a statistical technique used to model the relationship between one dependent variable and two or more independent variables
- It extends simple linear regression by allowing multiple predictors to influence the outcome

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 \ldots + w_n x_n + \varepsilon$$

$\hat{y}$: Dependent variable (response)

$x_1, x_2, \ldots, x_n$: Independent variable

ε: noise

w_0:intercept

$w_1, w_2, \ldots w_n$: coefficients

Example 2

x_1	x_2	y
1	2	2.2
2	1	2.8
3	3	4.5
4	4	6.3
5	5	8.1

$$E(w_0, w_1, w_2) = \frac{1}{2m}\sum_{i=1}^{m}(w_0 + w_1x_{1i} + w_2x_{2i} - y_i)^2$$

$$\frac{\delta E(W)}{\delta w_0} = \frac{1}{m}\sum_{i=1}^{m}(w_0+w_1x_{1i} + w_2x_{2i} - y_i)$$

$$\frac{\delta E(W)}{\delta w_1} = \frac{1}{m}\sum_{i=1}^{m}(w_0+w_1x_{1i} + w_2x_{2i} - y_i)\, x_{1i}$$

$$\frac{\delta E(W)}{\delta w_2} = \frac{1}{m}\sum_{i=1}^{m}(w_0+w_1x_{1i} + w_2x_{2i} - y_i)\, x_{2i}$$

$$w_0 = 0$$
$$w_1 = 0$$
$$w_2 = 0$$

Iteration 1

$$\frac{\delta E(W)}{\delta w_0} = \frac{(-2.2 - 2.8 - 4.5 - 6.3 - 8.1)}{5} = -4.78$$

$$\frac{\delta E(W)}{\delta w_1} = \frac{(-2.2 \times 1 - 2.8 \times 2 - 4.5 \times 3 - 6.3 \times 4 - 8.1 \times 5)}{5} = -17.4$$

$$\frac{\delta E(W)}{\delta w_2} = \frac{(-2.2 \times 2 - 2.8 \times 1 - 4.5 \times 3 - 6.3 \times 4 - 8.1 \times 5)}{5} = -17.28$$

$$w_0 = 0 - 0.1 \times -4.78 = 0.478$$

$$w_1 = 0 - 0.1 \times -17.4 = 1.74$$

$$w_2 = 0 - 0.1 \times -17.28 = 1.728$$

x_1	x_2	y	$\hat{y}$	Error	w_0	w_1	w_2
1	2	2.2	0	-2.2	0.478		
2	1	2.8	0	-2.8			
3	3	4.5	0	-4.5			
4	4	6.3	0	-6.3			
5	5	8.1	0	-8.1			

Iteration 2

$$w_0 = 0.847$$

$$w_1 = 3.088$$

$$w_2 = 3.064$$

Final

- $Y = 0.043 + 1.089X_1 + 0.489X_2$