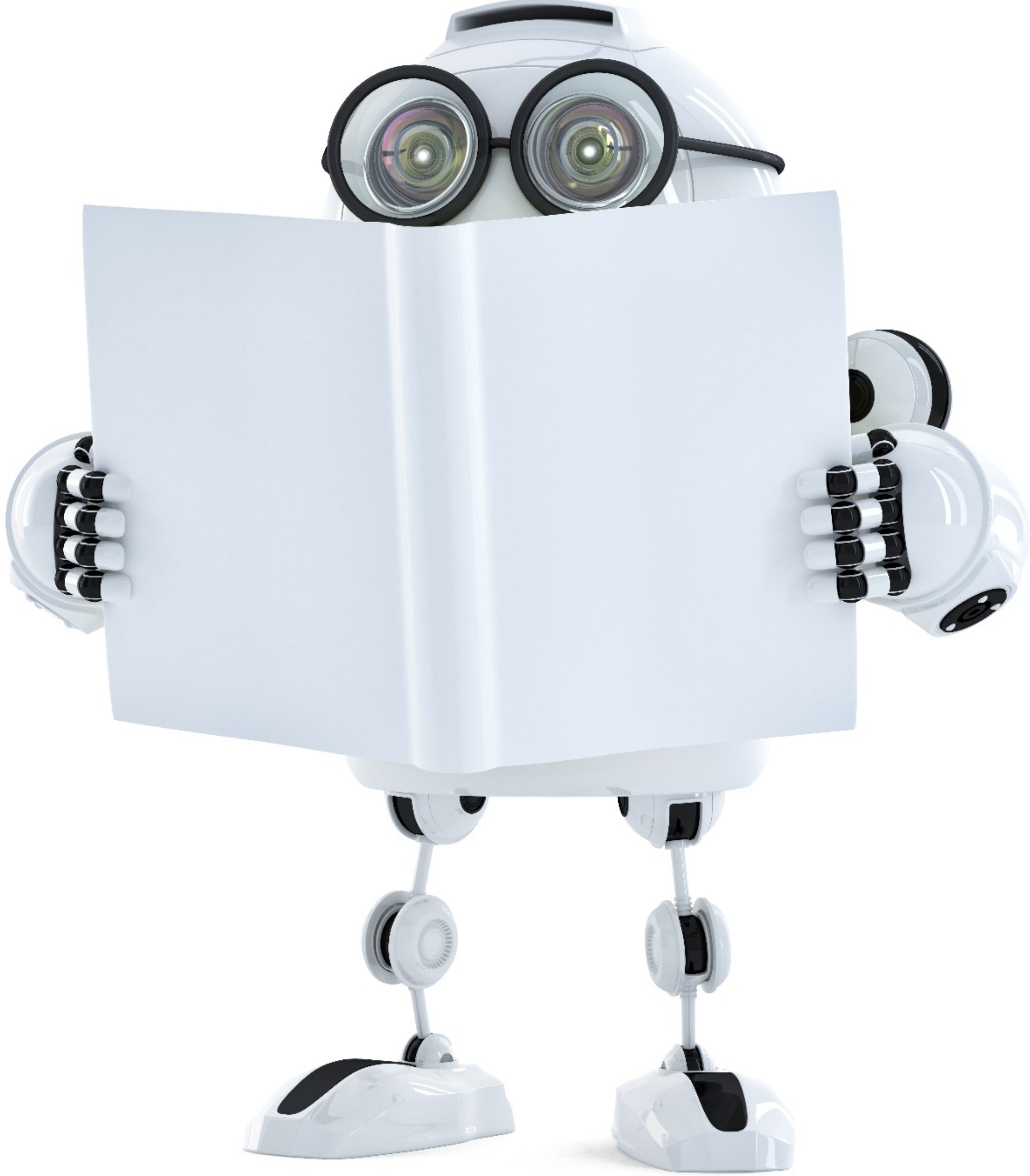
Normal equation

Linear Regression



Director of TEAMLAB
Sungchul Choi

Normal equation

Cost Function을 최소화하는 방법

$$\arg\min_{\theta} \ \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$y^{(1)} = w_0 + w_1 x^{(1)} + \epsilon^{(1)}$$
$$y^{(2)} = w_0 + w_1 x^{(2)} + \epsilon^{(2)}$$
$$y^{(3)} = w_0 + w_1 x^{(3)} + \epsilon^{(3)}$$
$$y^{(4)} = w_0 + w_1 x^{(4)} + \epsilon^{(4)}$$
$$y^{(5)} = w_0 + w_1 x^{(5)} + \epsilon^{(5)}$$

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ 1 & x^{(4)} \\ 1 & x^{(5)} \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \qquad \mathbf{y} = \mathbf{X}\mathbf{w}$$

$$\mathbf{y} = \mathbf{X}\mathbf{w}$$

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ 1 & x^{(4)} \\ 1 & x^{(5)} \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$J = \frac{1}{2}\sum_{i=1}^{m}(w_1 x^{(i)} + w_0 - y^{(i)})^2$$

$$\frac{\partial J}{\partial w_0} = \sum (w_1 x^{(i)} + w_0 - y^{(i)}) = 0$$

$$\frac{\partial J}{\partial w_1} = \sum (w_1 x^{(i)} + w_0 - y^{(i)}) x^{(i)} = 0$$

위 식을 만족하는 $\hat{w}_j$ 값을 구하기

$$\hat{w}_0 m + \hat{w}_1 \sum x^{(i)} = \sum y^{(i)}$$

$$\hat{w}_0 \sum x^{(i)} + \hat{w}_1 \sum (x^{(i)})^2 = \sum y^{(i)} x^{(i)}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} m & \sum x^{(i)} \\ \sum x^{(i)} & \sum (x^{(i)})^2 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})\hat{\mathbf{w}} = \mathbf{X}^T\mathbf{y}$$

$$\Downarrow$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ 1 & x^{(4)} \\ 1 & x^{(5)} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{bmatrix}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} m & \sum x^{(i)} \\ \sum x^{(i)} & \sum (x^{(i)})^2 \end{bmatrix} = \begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum (x^{(i)})^2 \end{bmatrix}$$

$$\begin{aligned} |\mathbf{X}^T\mathbf{X}| &= m\sum(x^{(i)})^2 - (m\bar{x})^2 \\ &= m(\sum(x^{(i)})^2 - m\bar{x}^2) \\ &= m\sum(x^{(i)} - \bar{x})^2 \end{aligned}$$

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$$

$$\begin{aligned} \sum(X_i - \mu)^2 &= \sum(X_i^2 - 2X_i\mu + \mu^2) \\ &= \sum X_i^2 - \sum 2X_i\mu + \sum \mu^2 \\ &= \sum X_i^2 - 2\mu\sum X_i + N\mu^2 \\ &= \sum X_i^2 - 2\mu N\mu_i + N\mu^2 \\ &= \sum X_i^2 - 2N\mu^2 + N\mu^2 \\ &= \sum X_i^2 - N\mu^2 \end{aligned}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} m & \sum x^{(i)} \\ \sum x^{(i)} & \sum (x^{(i)})^2 \end{bmatrix} = \begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum (x^{(i)})^2 \end{bmatrix}$$

$$\begin{aligned}(\mathbf{X}^T\mathbf{X})^{-1} &= \frac{1}{m\sum(x^{(i)} - \bar{x})^2} \begin{bmatrix} \sum (x^{(i)})^2 & -m\bar{x} \\ -m\bar{x} & m \end{bmatrix} \\ &= \frac{1}{\sum(x^{(i)} - \bar{x})^2} \begin{bmatrix} \sum (x^{(i)})^2/m & -\bar{x} \\ -\bar{x} & m \end{bmatrix}\end{aligned}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{m\sum(x^{(i)} - \bar{x})^2}\begin{bmatrix} \sum(x^{(i)})^2 & -m\bar{x} \\ -m\bar{x} & m \end{bmatrix}$$

$$= \frac{1}{\sum(x^{(i)} - \bar{x})^2}\begin{bmatrix} \sum(x^{(i)})^2/m & -\bar{x} \\ -\bar{x} & m \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} \sum y^{(i)} \\ \sum x^{(i)}y^{(i)} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ 1 & x^{(4)} \\ 1 & x^{(5)} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{bmatrix}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix}$$

$$= \frac{1}{\sum(x^{(i)} - \bar{x})^2} \begin{bmatrix} \sum(x^{(i)})^2/m & -\bar{x} \\ -\bar{x} & m \end{bmatrix} \begin{bmatrix} \sum y^{(i)} \\ \sum x^{(i)} y^{(i)} \end{bmatrix}$$

$$\hat{w}_1 = \frac{\sum x^{(i)} y^{(i)} - m\bar{x}\bar{y}}{\sum(x^{(i)} - \bar{x})^2}$$

$$\hat{w}_0 = \bar{y} - \hat{w}_1\bar{x}$$

여러개의 변수 일 경우?

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} m & \sum x^{(i)} \\ \sum x^{(i)} & \sum (x^{(i)})^2 \end{bmatrix}$$

가 확대됨

결론

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Normal equation

- $\mathrm{X}^T\mathrm{X}$의 역행렬이 존재할 때 사용
- Iteration 등 사용자 지정 parameter가 없음
- Feature가 많으면 계산 속도가 느려짐

TEAMLAB

Human knowledge belongs to the world.