모두의 딥러닝 (Deep learning)

Nambeom Kim (nbumkim@gmail.com)

Least square estimation (LSE)

선형회귀

- 지도학습
- 목적변수 (반응변수)가 연속형인 경우
- 정규성, 독립성, 등분산성을 만족해야 함
- Feature 가 하나인 경우 단순회귀 (Simple), Feature 가 여러 개 인 경우 중회귀 (Multiple)
- 2차항 이상이 포함된 경우 다항회귀 (Polynomial regression)

단순회귀 (Simple linear regression)

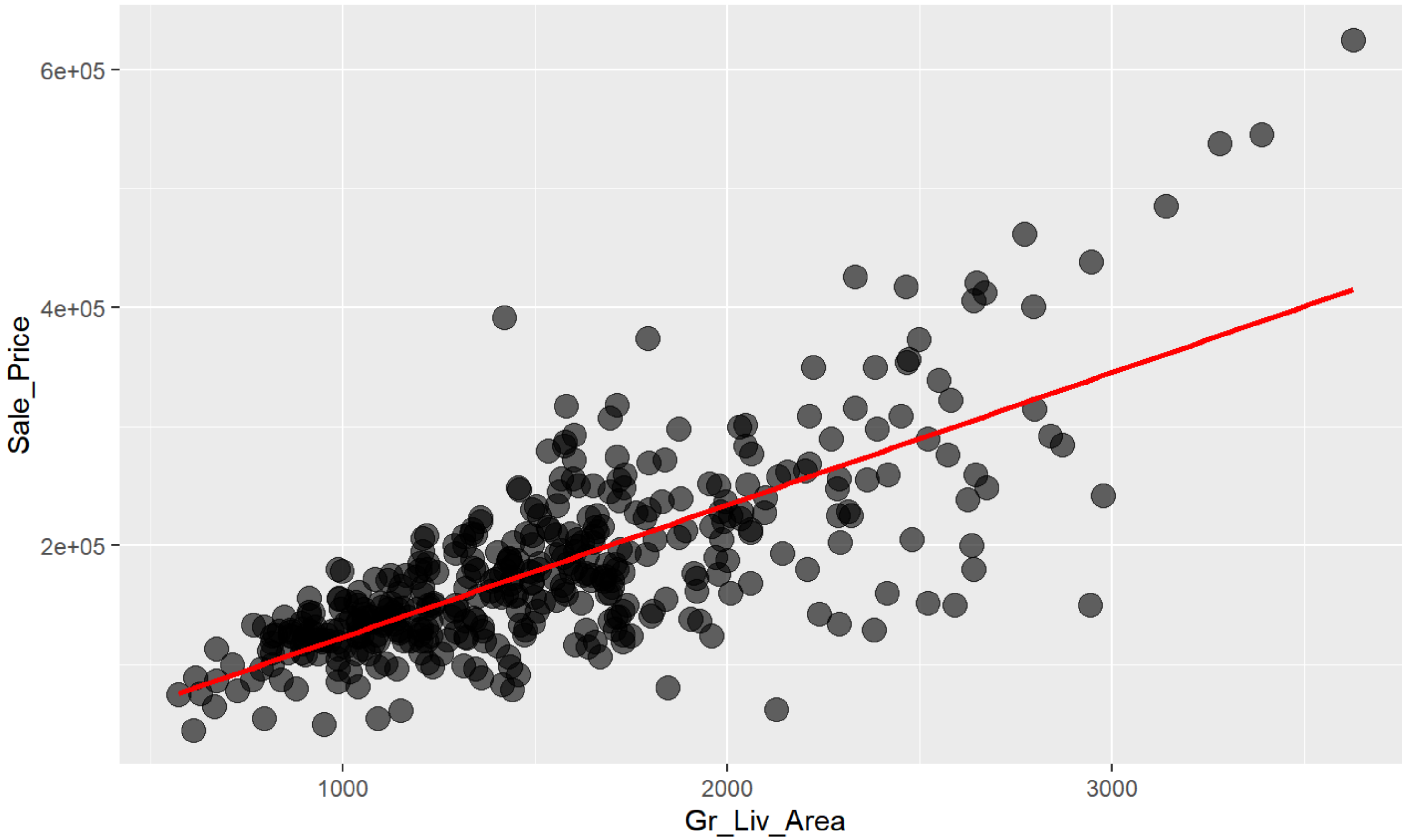
$$f(x_i) = Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \;\; for \; i = 1, 2, \ldots, n$$

- $\epsilon_i \sim^{i.i.d} N(0, \sigma^2)$, $\forall i$
- $Independency$, $Normality$, $Homoscedasticity$

```
head(ames[, c("Sale_Price", "Gr_Liv_Area")], n = 10)
```

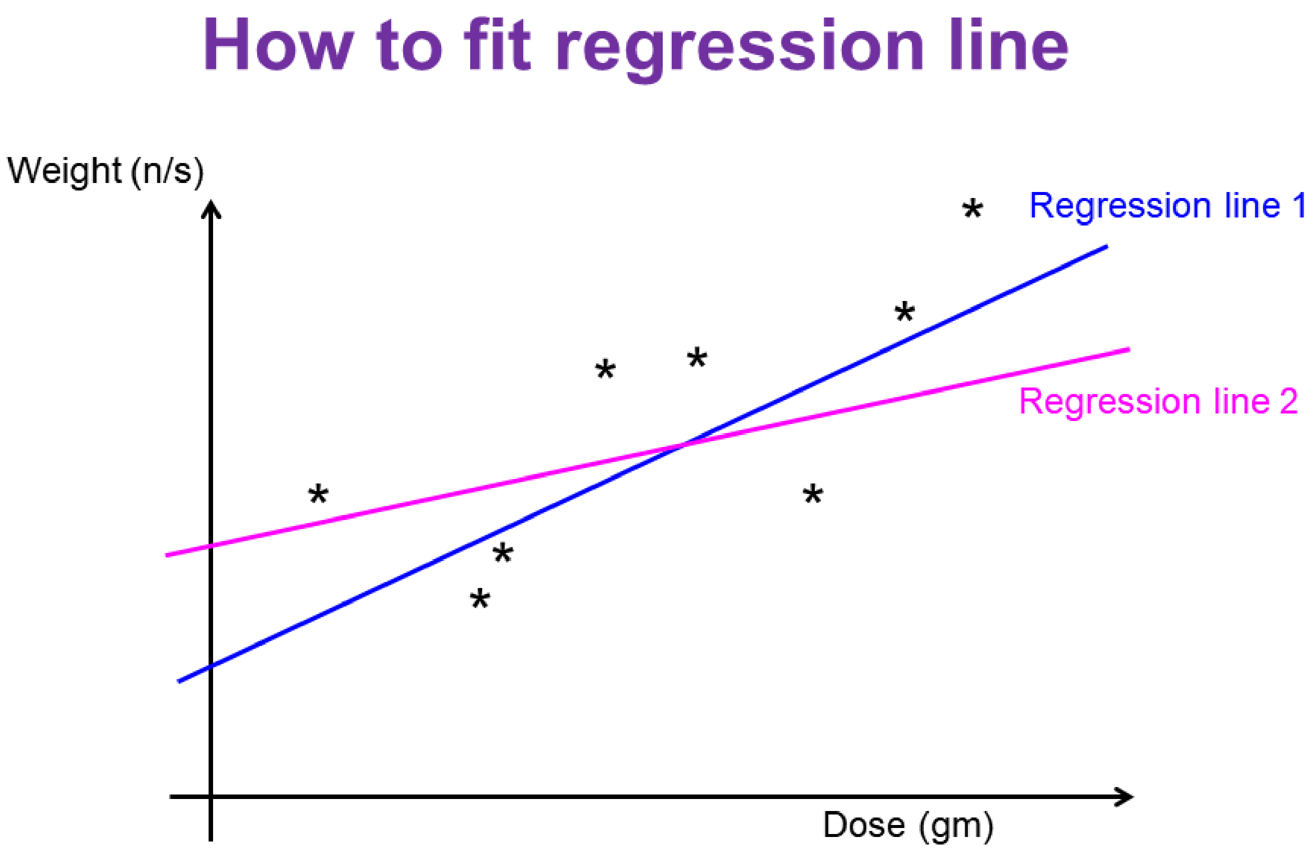
```
# A tibble: 10 x 2
   Sale_Price Gr_Liv_Area
        <int>       <int>
 1     215000        1656
 2     105000         896
 3     172000        1329
 4     244000        2110
 5     189900        1629
 6     195500        1604
 7     213500        1338
 8     191500        1280
 9     236500        1616
10     189000        1804
```

Gr_Liv_Area 와 Sale_Price 간의 관계



최소제곱추정량 (Ordinary least square estimation)

- Linear model, $\hat{y}(x_i) = \hat{\theta}_0 + \hat{\theta}_1 x_i$



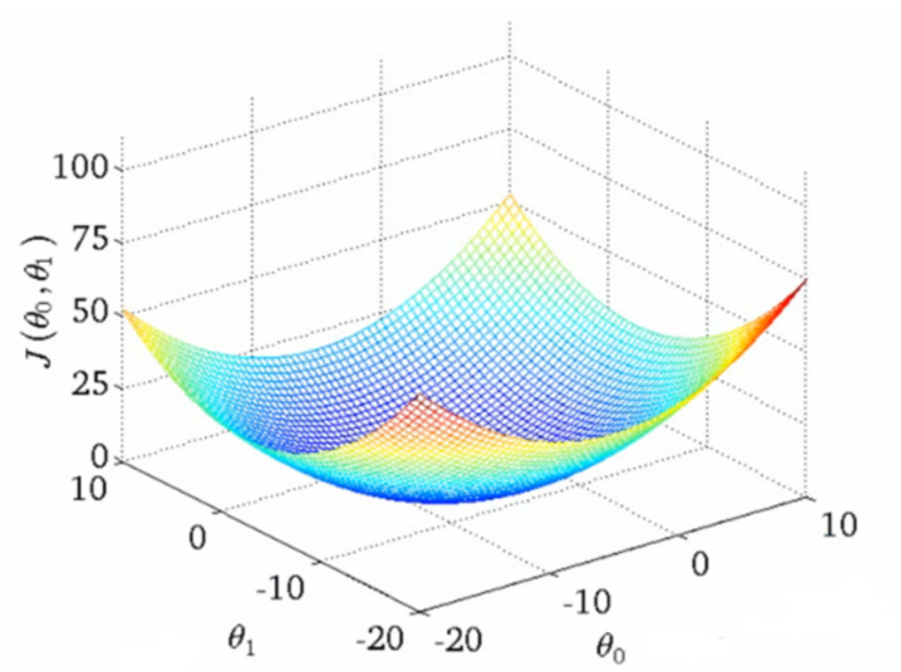
손실함수 (Loss function, J)

$J(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\theta}_0 - \hat{\theta}_1 x)^2 = \sum_{i=1}^{n} e_i^2$
- $\hat{y}_i = \hat{\theta}_0 + \hat{\theta}_1 x_i$
- Sum of Squares of the Errors (SSE) = $\sum_i^n e_i^2$

- Goal is to solve for $\hat{\theta}_0$ and $\hat{\theta}_1$ to minimize the objective function.

$$\hat{\theta}_o, \hat{\theta}_1 = argmin_{\theta_0,\theta_1} \sum_i^n e_i^2 = argmin_{\theta_0,\theta_1} J(\theta)$$

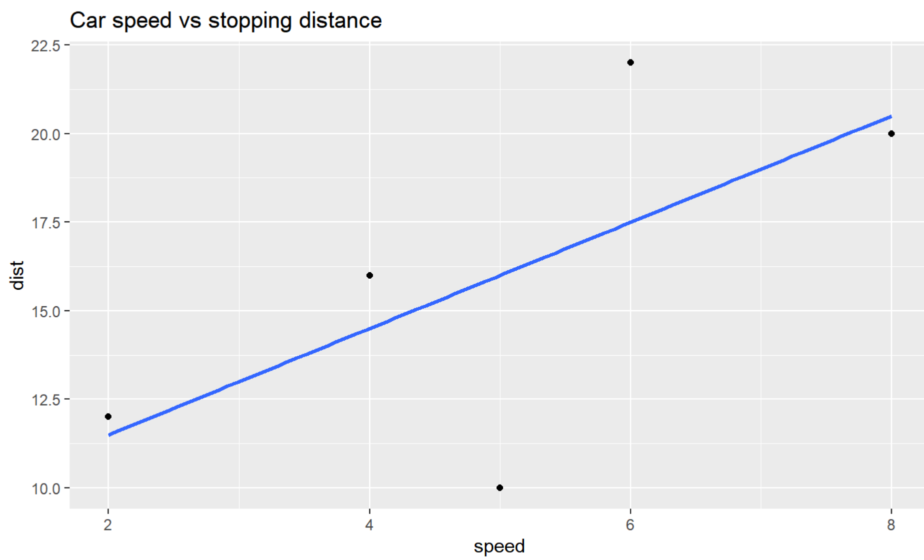
Convex function



- $\dfrac{\partial J(\theta)}{\partial \hat{\theta}_0} = 0, \quad \hat{\theta}_0 = \overline{y} - \hat{\theta}_1 \overline{x}$
- $\dfrac{\partial J(\theta)}{\partial \hat{\theta}_1} = 0, \quad \hat{\theta}_1 = \dfrac{\sum_i^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_i^n (x_i - \overline{x})^2}$
- $\sum_i^n (x_i - \overline{x})(y_i - \overline{y}) = S_{xy}$
- $\sum_i^n (x_i - \overline{x})^2 = S_{xx}$

자동차 속력과 제동거리 간의 관계

```
# A tibble: 5 x 4
  speed speed_mean  dist dist_mean
  <dbl>      <dbl> <dbl>     <dbl>
1     2          5    12        16
2     4          5    16        16
3     5          5    10        16
4     8          5    20        16
5     6          5    22        16
```



중회귀 (Multiple linear regression)

$$f(x_i) = Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \;\; for\; i = 1, 2, \dots, n$$

- $\epsilon_i \sim^{i.i.d} N(\beta^T X, \sigma^2), \forall i$ - $\beta^T X = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$

```r
head(ames[, c("Sale_Price", "Gr_Liv_Area", "Year_Built")], n = 6)
```

```
# A tibble: 6 x 3
  Sale_Price Gr_Liv_Area Year_Built
       <int>       <int>      <int>
1     215000        1656       1960
2     105000         896       1961
3     172000        1329       1958
4     244000        2110       1968
5     189900        1629       1997
6     195500        1604       1998
```

$\boxed{\text{Multiple Regression}}$ Model in Matrix Form

- In matrix notation the multiple regression model is: **Y=Xβ + ε**

  where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad , \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad , \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_\kappa \end{bmatrix} \quad , \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{nk} \end{bmatrix}$$

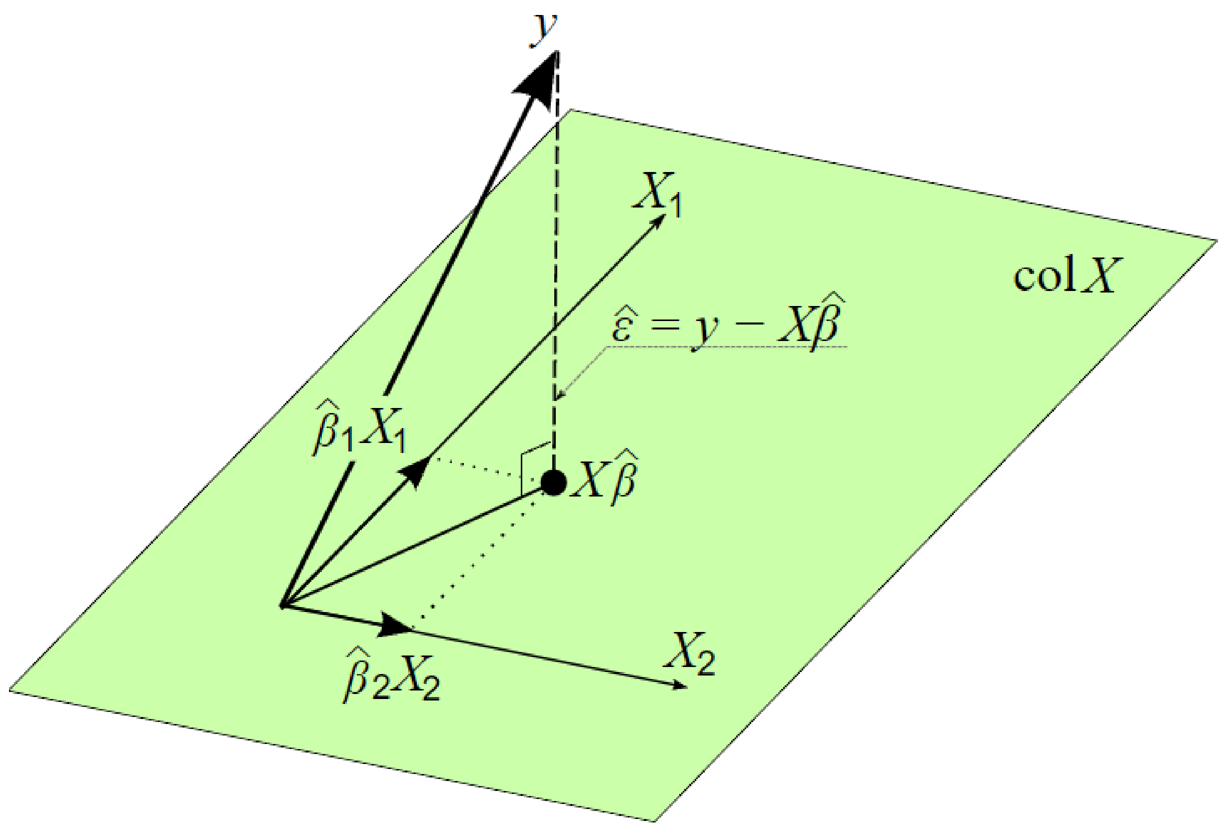
- Note, **Y** and **ε** are $n \times 1$ vectors, **β** is a $(k+1) \times 1$ vector and **X** is a $n \times (k+1)$ matrix.
- The Gauss-Markov assumptions are: $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$.
- These result in $E(\mathbf{Y}) = \mathbf{0}$, $\mathrm{Var}(\mathbf{Y}) = \sigma^2 \mathbf{I}$.
- The Least-Square estimate of **β** is $\mathbf{b} = \boxed{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}}$.

https://www.slideserve.com/verne/analysis-of-variance-in-matrix-form

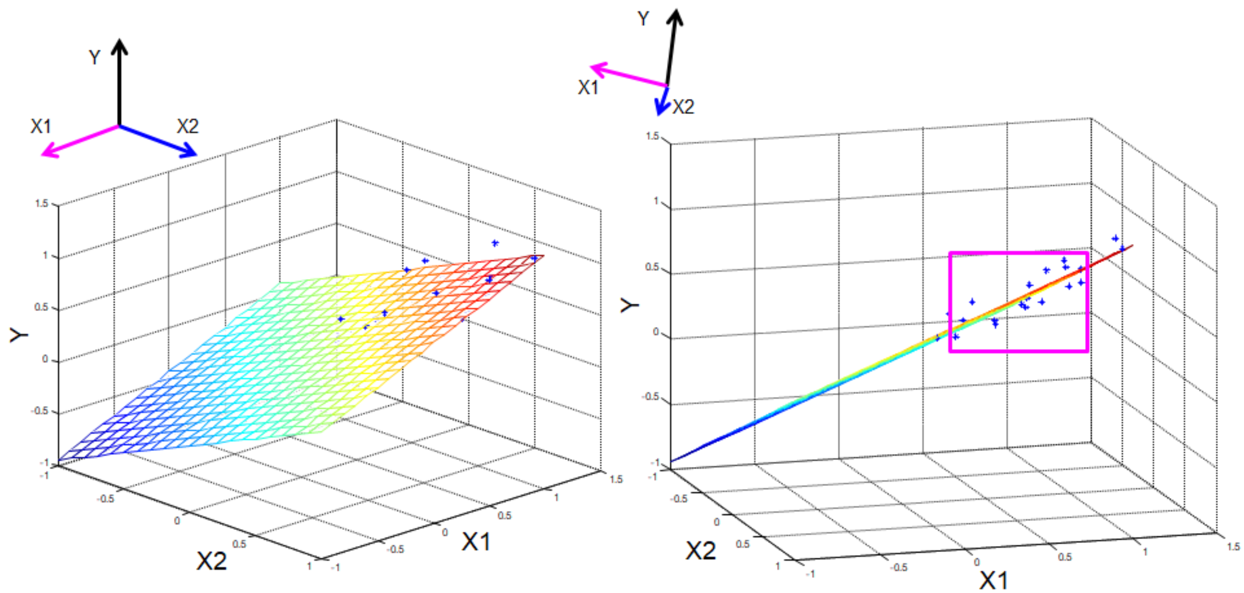
$\boxed{\text{OLS}}$ of Multiple linear regression

$\hat{\beta} = \boxed{(X'X)X^{-1}y}$
where, X = Feature matrix
y = Target vector

OLS geometric interpretation



2D regression plane



다항회귀 (Polynomial regression)

- $y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p + \epsilon_i$