Exercise 1

TA: David Lee
Email:
david20010603@alum.ccu.edu.tw

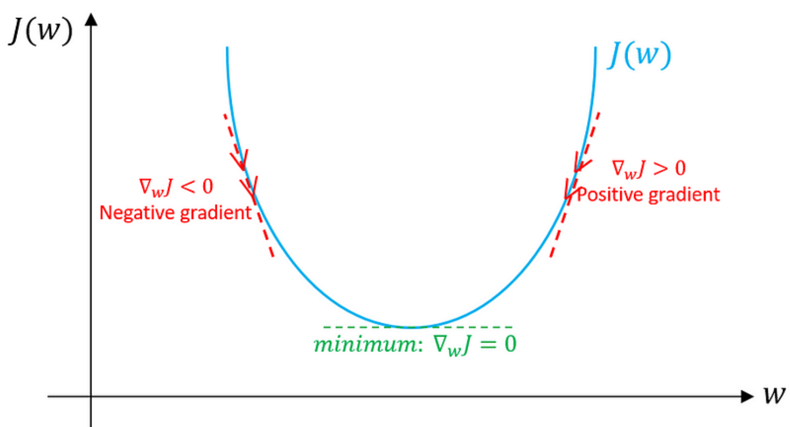
Understanding Linear Regression

- • Models relationship between input x and output y.
- • Types:
-   - Simple Linear Regression: $f(x) = wx$
-   - Linear Regression with Bias: $f(x) = w_1x + w_0$
- • Goal: Find the best parameters w or [w0, w1] that minimize the error.

The Meaning of Derivative(w)

- In linear regression, Derivative represents the rate of change of a function at a given point, also known as the gradient (slope).

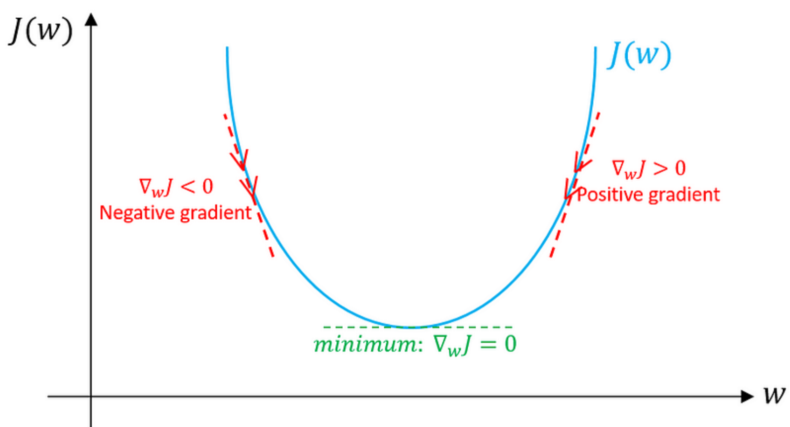
- In mathematics, the extremum of a function occurs at a critical point(臨界點), where the derivative is equal to zero.
- Therefore, we find the extremum by setting the rate of change of the function to zero.



The Meaning of Derivative(導數)

- 在線性回歸中，導數代表函數在某一點的變化率，也被稱為梯度（斜率）。

- 在**數學**中，函數的極值(大or小)發生在**臨界點**，即導數等於零的地方。
- 因此，我們透過設函數的變化率為 0 來求極值。



(a) Solving for w in $f(x) = wx$

- • Objective: Minimize the error

$$J(w) = \sum (wx_i - y_i)^2$$

- Find the first derivative of the function with respect to w(對函數J(w)做w偏微分)

$$\frac{\partial \mathrm{J}(w)}{\partial w} = \underbrace{\frac{\partial \mathrm{J}(w)}{\partial (wx_i - y_i)} \cdot \frac{\partial (wx_i - y_i)}{\partial w}}_{\text{Chain Rule}} = \frac{1}{n}\sum_{i=1}^{n} 2(wx_i - y_i) \cdot x_i = \frac{2}{n}\sum_{i=1}^{n} x_i \cdot (wx_i - y_i)$$

$$\cancel{\frac{2}{n}}\sum_{i=1}^{n} x_i \cdot (wx_i - y_i) = 0$$

$$\sum_{i=1}^{n} x_i\, y_i = \sum_{i=1}^{n} w{x_i}^2$$

$$\sum_{i=1}^{n} \frac{x_i y_i}{{x_i}^2} = w$$

- • Solution:

$$w = \sum \frac{x_i y_i}{{x_i}^2}$$

$$w = 1.3$$

(b) Solving for w0, w1 using Normal Equation

- • Model: $f(x_i) = y_i = w_1 x_i + w_0$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_i \end{bmatrix}_{i\times 1} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_i \end{bmatrix}_{i\times 2} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}_{2\times 1}$$

- •Apply Normal Equation: $w = (X^T X)^{-1} X^T Y$
- Matrix Formulation:

$$X = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, Y = \begin{bmatrix} -4 \\ 1 \\ 1 \\ 2 \\ 6 \end{bmatrix}$$

- Solution:

$$w = \frac{1}{146} \begin{bmatrix} 30 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ 39 \end{bmatrix} = \frac{1}{146} \begin{bmatrix} 102 \\ 183 \end{bmatrix}$$

(c) Predictions for x = 1

- • Using $f(x) = 1.3x$:Substitute x = 1, compute y.
- • Using $f(x) = 1.253x + 0.698$:Substitute x = 1, compute y.
- Solution:

$$(a)\ f(1) = 1.3$$
$$(b) f(1) = 1.253 + 0.698 = 1.951$$

(d) Gradient Descent Explanation

- • $J(w) = \frac{1}{n}\sum_{i=1}^{n}(wx_i - y_i)^2$
- • Update Rule:

$$W^{(t+1)} = w^t - \alpha\frac{\partial J}{\partial w}$$

$$\frac{\partial \mathrm{J}(w)}{\partial w} = \frac{2}{n}\sum_{i=1}^{n} x_i \cdot (wx_i - y_i)$$

- •Hyperparameters:
- \- Learning Rate: α = 0.1
- \- Initial Value: w = 0

(d) Gradient Descent Iterations

- • Iteration 1: Compute new w.
- • Iteration 2: Compute next w.
- • Show step-by-step calculations.

$$w_1 = w_0 - \frac{0.1}{5} \cdot [\sum_{i=1}^{n} x_i \cdot (w_0 x_i - y_i)]$$

$$w_1 = w_0 - \frac{0.1}{5} \cdot [(-3) \cdot 4 + (-1) \cdot (-1) + 0 \cdot (-1) + 2 \cdot (-2) + 4 \cdot (-6)] = -\frac{0.1}{5} \cdot -39$$
$$= 0.78$$

$$w_2 = w_1 - \frac{0.1}{5} \cdot [\sum_{i=1}^{n} x_i \cdot (0.78 x_i - y_i)]$$
$$w_2 = w_1 - \frac{0.1}{5} \cdot [-3 \cdot 1.66 + 1.78 + 2.044 + 4.288] = 0.78 - \frac{0.1}{5} \cdot (15.6) = 1.092$$

Problem 2

(a)

•

$$Linear\ regression\ equations: y_1 = w_0 x_1 + w_1 x_2, y_2 = w'_0 x_1 + w'_1 x_2$$

$$In\ matrix\ format: f(x) = Y = Xw$$

$$[y_1 \quad y_2] = [x_1 \quad x_2]\begin{bmatrix} w_0 & w'_0 \\ w_1 & w'_1 \end{bmatrix}$$

$$Y = \begin{bmatrix} 7 & 10 \\ 4 & 5 \\ 2 & 9 \\ 6 & 6 \end{bmatrix}_{4\times 2} \quad X = \begin{bmatrix} 3 & 5 \\ 2 & 2 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}_{4\times 2} \quad W = \begin{bmatrix} w_0 & w'_0 \\ w_1 & w'_1 \end{bmatrix}$$

$$To\ estimate\ W, use\ normal\ equation: W = (X^T X)^{-1}\, X^T Y = \begin{bmatrix} \frac{18}{11} & -1 \\ \frac{25}{26} & 3 \end{bmatrix}$$

$$y_1 = \frac{18}{11} x_1 + \frac{25}{26} x_2$$

$$y_2 = -1x_1 + 3x_2$$

(b)

- Substitute $x_1 = 7, x_2 = 4, into$
  $y_1 = w_0 x_1 + w_1 x_2, y_2 = w'_0 x_1 + w'_1 x_2$
- $To\ get\ the\ predicted\ outputs$

$$[y_1 \quad y_2] = [7 \quad 4]\begin{bmatrix} \frac{18}{11} & -1 \\ \frac{25}{66} & 3 \end{bmatrix} = \begin{bmatrix} \frac{428}{33} & 5 \end{bmatrix}$$