Understanding the Derivative in Linear Regression

https://www.youtube.com/watch?v=Ijd1PLmThsU

1. Problem Statement

We want to find the values of **β_0** (intercept) and **β_1** (slope) that minimize the **Mean Squared Error (MSE)** in Simple Linear Regression. The model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where:

- Y_i is the actual value.
- X_i is the independent variable.
- β_0, β_1 are parameters to be estimated.
- ϵ_i is the error term.

What is β_0 and β_1

1. **β_0 (Intercept)**:
    - It represents the value of **Y** when **X = 0**.
    - In other words, it is the predicted value of the dependent variable when the independent variable is zero.
    - Mathematically:

    $$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

    - Example: If you are predicting house prices based on square footage, β_0 would represent the price of a house with **zero** square footage (which may not always be meaningful but is necessary for the equation).
2. **β_1 (Slope)**:
    - It represents the **rate of change** of **Y** with respect to **X**.

- In simple terms, it shows how much **Y** increases (or decreases) when **X** increases by **one unit**.
- Mathematically, it is calculated as:

$\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$

- Example: If β_1 = 2000 in a house price model, it means **for every additional square foot, the price increases by $2000**.

Together, β_0 and β_1 define the **best-fit line** in simple linear regression:

$Y = \beta_0 + \beta_1 X + \epsilon$

---

2. The Cost Function (MSE)

To find the best β_0 and β_1, we minimize the **Mean Squared Error (MSE)**:

$$J(\beta_0, \beta_1) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

Substituting $\hat{Y}_i = \beta_0 + \beta_1 X_i$:

$$J(\beta_0, \beta_1) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i)^2$$

We minimize $J(\beta_0, \beta_1)$ by taking **partial derivatives** with respect to β_0 and β_1, then setting them to zero.

---

3. Derivative with Respect to β_0 (Intercept)

$$\frac{\partial J}{\partial \beta_0} = \frac{1}{N} \sum_{i=1}^{N} 2(Y_i - \beta_0 - \beta_1 X_i)(-1) = 0$$

Rearrange:

$\sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) = 0$

Dividing by N:

$\frac{1}{N} \sum_{i=1}^{N} Y_i - \beta_0 - \beta_1 \frac{1}{N} \sum_{i=1}^{N} X_i = 0$

Solving for β_0:

$\beta_0 = \bar{Y} - \beta_1 \bar{X}$

where $\bar{Y}$ and $\bar{X}$ are the means of Y and X respectively.

---

4. Derivative with Respect to β_1 (Slope)

$$\frac{\partial J}{\partial \beta_1} = \frac{1}{N} \sum_{i=1}^{N} 2(Y_i - \beta_0 - \beta_1 X_i)(-X_i) = 0$$

Expanding:

$\sum_{i=1}^{N} X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$

Substituting $\beta_0 = \bar{Y} - \beta_1 \bar{X}$:

$\sum_{i=1}^{N} X_i (Y_i - \bar{Y} + \beta_1 \bar{X} - \beta_1 X_i) = 0$

Rearrange:

$$\sum_{i=1}^{N} X_i Y_i - \bar{Y} \sum_{i=1}^{N} X_i + \beta_1 \bar{X} \sum_{i=1}^{N} X_i - \beta_1 \sum_{i=1}^{N} X_i^2 = 0$$

Solving for β_1:

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

---

5. Final Formulas

a) **Slope (β_1)**:

$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

b) **Intercept (β0\beta_0)**:

$\beta_0 = \bar{Y} - \beta_1 \bar{X}$

These are the **Normal Equations** for **Ordinary Least Squares (OLS) regression**.

---

6. Why Do We Do This?

- We need to minimize the error in our regression model.
- Taking derivatives helps us find the optimal values of and .

  β_0

  β_1
- This method is used in **Machine Learning**, **Deep Learning**, and **Statistics**.

---

7. Key Takeaways

✅ **Linear regression finds the best fit line** by minimizing MSE.

✅ **Partial derivatives help find optimal parameters** (slope & intercept).

✅ **Gradient Descent follows a similar approach**, using derivatives to optimize models.
