# Machine Learning Engineer Course
# Day 10

- Scratch Linear Regression -

DIVE INTO CODE

Thursday May 13th, 2021

DIOP Mouhamed

# Agenda

1. Check-in
2. Please note & How to proceed
3. Quick Review
4. Linear Regression
5. Linear Regression Class of Scikit-learn
6. Assignment
7. Optimization & Sample Code
8. Check-out

# Check-in

**3 minutes** Please post the following point to Zoom chat.

**Q. What do you still find hard to understand in Machine Learning ?**

(Anything is fine. Even if you have already passed that part)

# 🚨 Please Note

**How to proceed with this course and precautions**
You will be the leader in the IT industry in Vietnam.

## ❶ Advance at top speed

We do not bottom up

## ❷ Promote autonomous self-propelled

We do not accept unexplained questions

## ❸ Focus on problem-solving ability

We do not give lectures on building up the foundation

# How to proceed – Objective (1/2)

## What is the purpose?

1. **Understanding Statistical Models**
   **Understanding Linear Regression through Scratch**

2. **Create a class**
   **Become familiar with object-oriented implementation**

3. **Learn about the computational process of machine learning from mathematical formulas**
   **To be able to put formulas into code**

# How to proceed – Objective (2/2)

**How to solve problems**
**“Scratch Linear Regression”**

【Problem1】 Hypothetical function
【Problem2】 Steepest descent
【Problem 3】 Estimated
【Problem 4】 Mean squared error
【Problem 5】 Objective function
【Problem 6】 Learning and estimation
【Problem 7】 Learning curve plot
【Problem 8】 (Advance task) Removal of bias term
【Problem 9】 (Advance task) Multidimensional feature quantity
【Problem 10】 (Advance task) Derivation of update formula
【Problem 11】 (Advance task) Problem of local optimum solution

# Quick Review (ML Scratch)

**1. What is it about ?**

**It's about creating a class/function and make it work the same as the one you would import from the libraries.**
**We can take the example of the train_test_split class of scikit-learn that is in the model_selection module which you will be implementing from scratch in this SPRINT.**

**2. Why ?**

- ✓ **Make it easier to understand the theory and mathematic formulas.**
- ✓ **Reducing ambiguity in using libraries**
- ✓ **Making existing implementations easier to read**
- ✓ **Improving your coding skills**
- ✓ **Deeper understanding of the algorithms**

# Target audience for this assignment

**1) Those who can write code to train and estimate using scikit-learn's linear regression model.**

**2) Those who know a little bit about gradient descent algorithm (1)**

**(1) The problem of descending Mt. Fuji in week 2 class assignment 2**

# Linear Regression – The Flow

## The Flow of the Linear Regression

(1) Formulate an equation (hypothetical function) to derive the predicted value

(2) Calculate the error between the target variable and the predicted value.

(3) Set up a problem to minimize this error and formulate an equation (objective function).

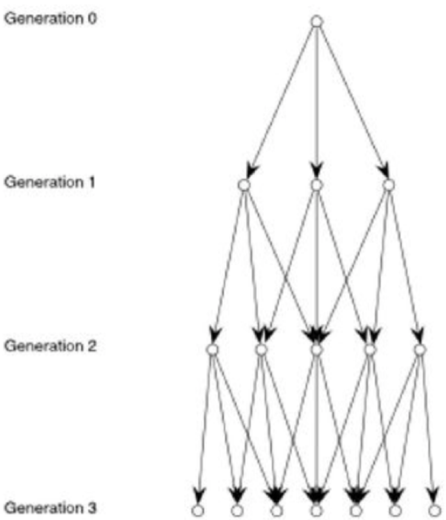
(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).

(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.

# What is linear regression?

## From History

The word originated in the 19th century, when Francis Gorton observed biological data and found that the height of tall parents did not necessarily pass straight to their offspring, but tended to regress back to the mean of the population, or "regression.



# What is linear regression?

In machine learning, learning a mapping from x to y is considered to be solving a regression problem (1).

Linear (2) regression is a model (approximate function) that shows how much the objective variable (y) depends on the explanatory variable (x).

If there is one explanatory variable, a linear simple regression model is used, and if there are two or more explanatory variables, a linear multiple regression model is used.

(1) y is a continuous variable in this case.
(2) Here are the conditions that the linear form satisfies. https://mathtrain.jp/linear

# What is linear regression?

## From History

Golton invented "linear regression" to analyze this event.

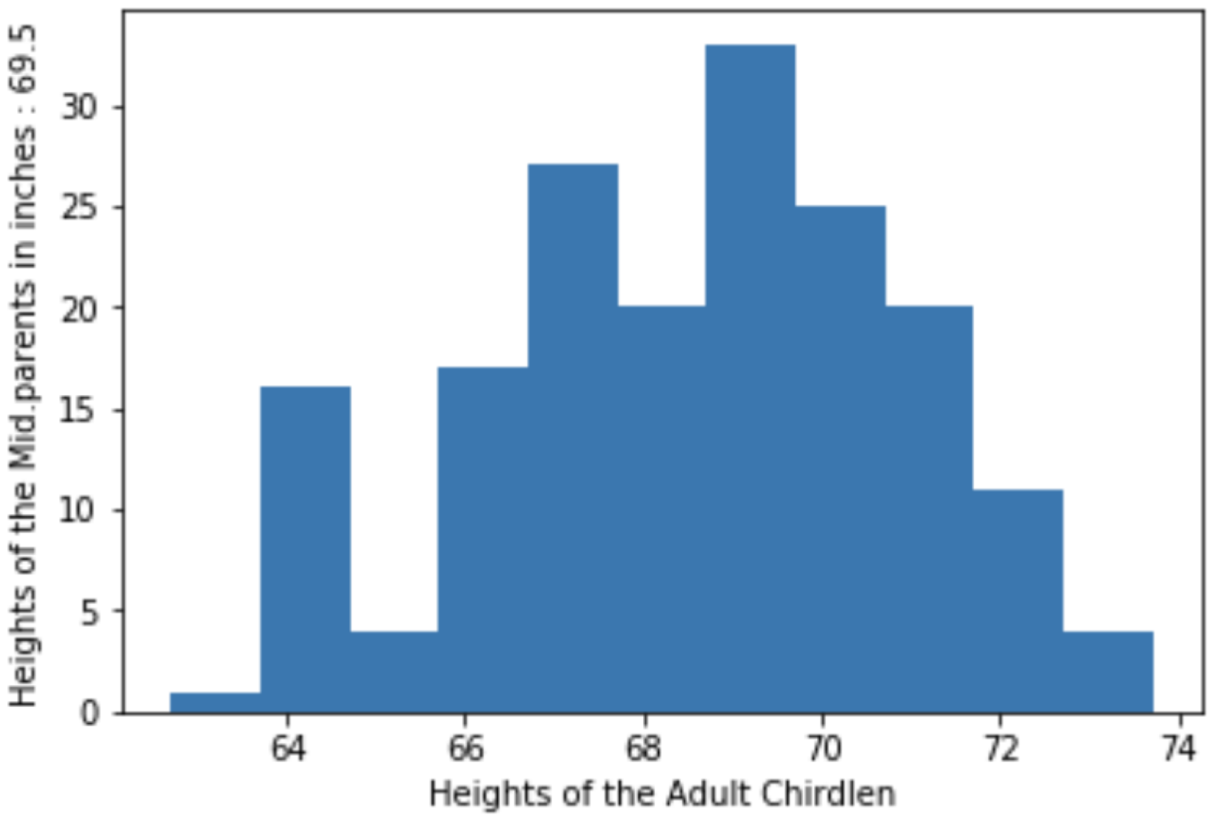
| Height of Mid-Parent (in inches)[a] | Height of Adult Child | | | | | | | | | | | | | | Total No. of Adult Children | Total No. of Mid-parents | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | < 61.7 | 62.2 | 63.2 | 64.2 | 65.2 | 66.2 | 67.2 | 68.2 | 69.2 | 70.2 | 71.2 | 72.2 | 73.2 | > 73.7 | | | |
| > 73.0 | — | — | — | — | — | — | — | — | — | — | — | 1 | 3 | — | 4 | 5 | — |
| 72.5 | — | — | — | — | — | — | — | 1 | 2 | 1 | 2 | 7 | 2 | 4 | 19 | 6 | 72.2 |
| 71.5 | — | — | — | — | 1 | 3 | 4 | 3 | 5 | 10 | 4 | 9 | 2 | 2 | 43 | 11 | 69.9 |
| 70.5 | 1 | — | 1 | — | 1 | 1 | 3 | 12 | 18 | 14 | 7 | 4 | 3 | 3 | 68 | 22 | 69.5 |
| 69.5 | — | — | 1 | 16 | 4 | 17 | 27 | 20 | 33 | 25 | 20 | 11 | 4 | 5 | 183 | 41 | 68.9 |
| 68.5 | 1 | — | 7 | 11 | 16 | 25 | 31 | 34 | 48 | 21 | 18 | 4 | 3 | — | 219 | 49 | 68.2 |
| 67.5 | — | 3 | 5 | 14 | 15 | 36 | 38 | 28 | 38 | 19 | 11 | 4 | — | — | 211 | 33 | 67.6 |
| 66.5 | — | 3 | 3 | 5 | 2 | 17 | 17 | 14 | 13 | 4 | — | — | — | — | 78 | 20 | 67.2 |
| 65.5 | 1 | — | 9 | 5 | 7 | 11 | 11 | 7 | 7 | 5 | 2 | 1 | — | — | 66 | 12 | 66.7 |
| 64.5 | 1 | 1 | 4 | 4 | 1 | 5 | 5 | — | 2 | — | — | — | — | — | 23 | 5 | 65.8 |
| < 64.0 | 1 | — | 2 | 4 | 1 | 2 | 2 | 1 | 1 | — | — | — | — | — | 14 | 1 | — |
| Totals | 5 | 7 | 32 | 59 | 48 | 117 | 138 | 120 | 167 | 99 | 64 | 41 | 17 | 14 | 928 | 205 | — |
| Medians | — | — | 66.3 | 67.8 | 67.9 | 67.7 | 67.9 | 68.3 | 68.5 | 69.0 | 69.0 | 70.0 | — | — | — | — | — |


https://slideplayer.com/slide/12965344/

# Creating a Histogram from a Frequency Distribution Table

## Mean height of parents: Frequency distribution table

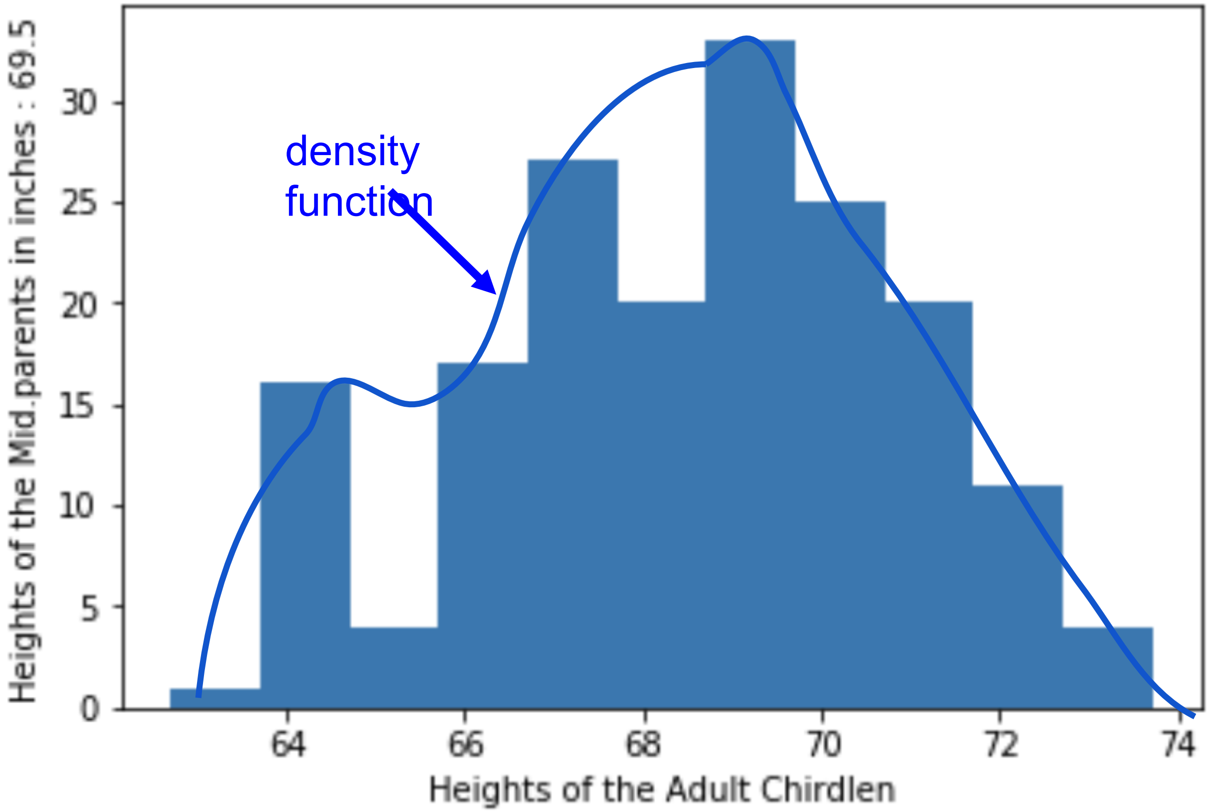
| Height of Mid-Parent (in inches)[a] | Height of Adult Child | | | | | | | | | | | | | | Total No. of Adult Children | Total No. of Mid-parents | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | < 61.7 | 62.2 | 63.2 | 64.2 | 65.2 | 66.2 | 67.2 | 68.2 | 69.2 | 70.2 | 71.2 | 72.2 | 73.2 | > 73.7 | | | |
| 69.5 | — | — | 1 | 16 | 4 | 17 | 27 | 20 | 33 | 25 | 20 | 11 | 4 | 5 | 183 | 41 | 68.9 |



Kernel density estimation (kde) also plots the density function.
It is a little skewed, but looks like a normal distribution.

## Mean height of parents: Frequency distribution table

| Height of Mid-Parent (in inches)[a] | Height of Adult Child | | | | | | | | | | | | | | Total No. of Adult Children | Total No. of Mid-parents | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | < 61.7 | 62.2 | 63.2 | 64.2 | 65.2 | 66.2 | 67.2 | 68.2 | 69.2 | 70.2 | 71.2 | 72.2 | 73.2 | > 73.7 | | | |
| 69.5 | — | — | 1 | 16 | 4 | 17 | 27 | 20 | 33 | 25 | 20 | 11 | 4 | 5 | 183 | 41 | 68.9 |



# Arrange the density function according to the frequency distribution table for each row.

## From each row showing the average of the parents' heights, we find a small normal distribution.

A series of normal distributions can be seen on the table data.

| Height of Mid-Parent (in inches)[a] | Height of Adult Child | | | | | | | | | | | | | | Total No. of Adult Children | Total No. of Mid-parents | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | < 61.7 | 62.2 | 63.2 | 64.2 | 65.2 | 66.2 | 67.2 | 68.2 | 69.2 | 70.2 | 71.2 | 72.2 | 73.2 | > 73.7 | | | |
| > 73.0 | — | — | — | — | — | — | — | — | — | — | — | 1 | 3 | — | 4 | 5 | — |
| 72.5 | — | — | — | — | — | — | — | 1 | 2 | 1 | 2 | 7 | 2 | 4 | 19 | 6 | 72.2 |
| 71.5 | — | — | — | — | 1 | 3 | 4 | 3 | 5 | 10 | 4 | 9 | 2 | 2 | 43 | 11 | 69.9 |
| 70.5 | 1 | — | 1 | — | 1 | 1 | 3 | 12 | 18 | 14 | 7 | 4 | 3 | 3 | 68 | 22 | 69.5 |
| 69.5 | — | — | 1 | 16 | 4 | 17 | 27 | 20 | 33 | 25 | 20 | 11 | 4 | 5 | 183 | 41 | 68.9 |
| 68.5 | 1 | — | 7 | 11 | 16 | 25 | 31 | 34 | 48 | 21 | 18 | 4 | 3 | — | 219 | 49 | 68.2 |
| 67.5 | — | 3 | 5 | 14 | 15 | 36 | 38 | 28 | 38 | 19 | 11 | 4 | — | — | 211 | 33 | 67.6 |
| 66.5 | — | 3 | 3 | 5 | 2 | 17 | 17 | 14 | 13 | 4 | — | — | — | — | 78 | 20 | 67.2 |
| 65.5 | 1 | — | 9 | 5 | 7 | 11 | 11 | 7 | 7 | 5 | 2 | 1 | — | — | 66 | 12 | 66.7 |
| 64.5 | 1 | 1 | 4 | 4 | 1 | 5 | 5 | — | 2 | — | — | — | — | — | 23 | 5 | 65.8 |
| < 64.0 | 1 | — | 2 | 4 | 1 | 2 | 2 | 1 | 1 | — | — | — | — | — | 14 | 1 | — |
| Totals | 5 | 7 | 32 | 59 | 48 | 117 | 138 | 120 | 167 | 99 | 64 | 41 | 17 | 14 | 928 | 205 | — |
| Medians | — | — | 66.3 | 67.8 | 67.9 | 67.7 | 67.9 | 68.3 | 68.5 | 69.0 | 69.0 | 70.0 | — | — | — | — | — |


https://slideplayer.com/slide/12965344/

# What is linear regression?

If we connect the means of these distributions, we can draw a straight line.

Golton invented "linear regression" to analyze this event.

| Height of Mid-Parent (in inches)[a] | Height of Adult Child | | | | | | | | | | | | | | Total No. of Adult Children | Total No. of Mid-parents | Median |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | < 61.7 | 62.2 | 63.2 | 64.2 | 65.2 | 66.2 | 67.2 | 68.2 | 69.2 | 70.2 | 71.2 | 72.2 | 73.2 | > 73.7 | | | |
| > 73.0 | — | — | — | — | — | — | — | — | — | — | — | 1 | 3 | — | 4 | 5 | — |
| 72.5 | — | — | — | — | — | — | — | 1 | 2 | 1 | 2 | 7 | 2 | 4 | 19 | 6 | 72.2 |
| 71.5 | — | — | — | — | 1 | 3 | 4 | 3 | 5 | 10 | 4 | 9 | 2 | 2 | 43 | 11 | 69.9 |
| 70.5 | 1 | — | 1 | — | 1 | 1 | 3 | 12 | 18 | 14 | 7 | 4 | 3 | 3 | 68 | 22 | 69.5 |
| 69.5 | — | — | 1 | 16 | 4 | 17 | 27 | 20 | 33 | 25 | 20 | 11 | 4 | 5 | 183 | 41 | 68.9 |
| 68.5 | 1 | — | 7 | 11 | 16 | 25 | 31 | 34 | 48 | 21 | 18 | 4 | 3 | — | 219 | 49 | 68.2 |
| 67.5 | — | 3 | 5 | 14 | 15 | 36 | 38 | 28 | 38 | 19 | 11 | 4 | — | — | 211 | 33 | 67.6 |
| 66.5 | — | 3 | 3 | 5 | 2 | 17 | 17 | 14 | 13 | 4 | — | — | — | — | 78 | 20 | 67.2 |
| 65.5 | 1 | — | 9 | 5 | 7 | 11 | 11 | 7 | 7 | 5 | 2 | 1 | — | — | 66 | 12 | 66.7 |
| 64.5 | 1 | 1 | 4 | 4 | 1 | 5 | 5 | — | 2 | — | — | — | — | — | 23 | 5 | 65.8 |
| < 64.0 | 1 | — | 2 | 4 | 1 | 2 | 2 | 1 | 1 | — | — | — | — | — | 14 | 1 | — |
| Totals | 5 | 7 | 32 | 59 | 48 | 117 | 138 | 120 | 167 | 99 | 64 | 41 | 17 | 14 | 928 | 205 | — |
| Medians | — | — | 66.3 | 67.8 | 67.9 | 67.7 | 67.9 | 68.3 | 68.5 | 69.0 | 69.0 | 70.0 | — | — | — | — | — |

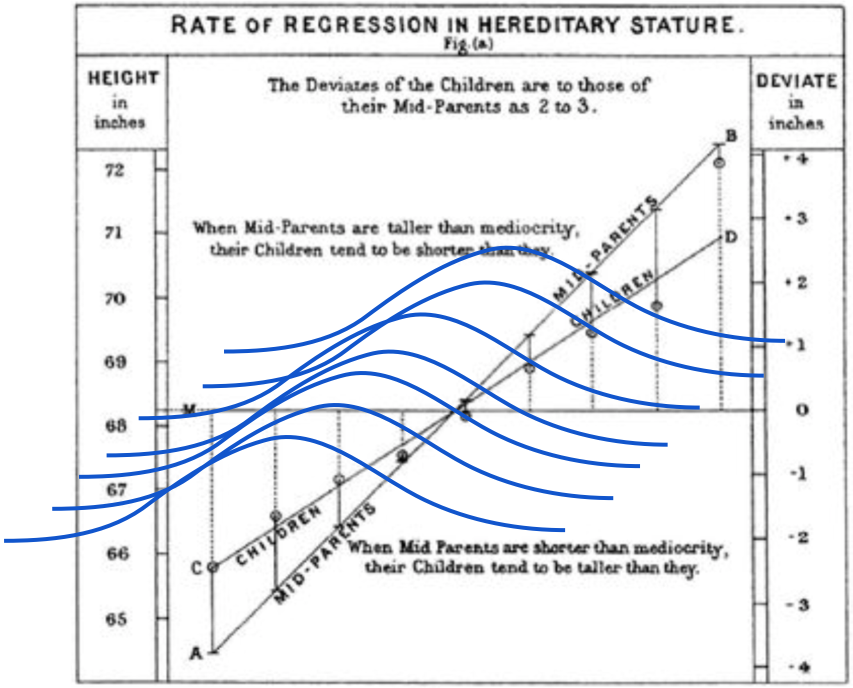


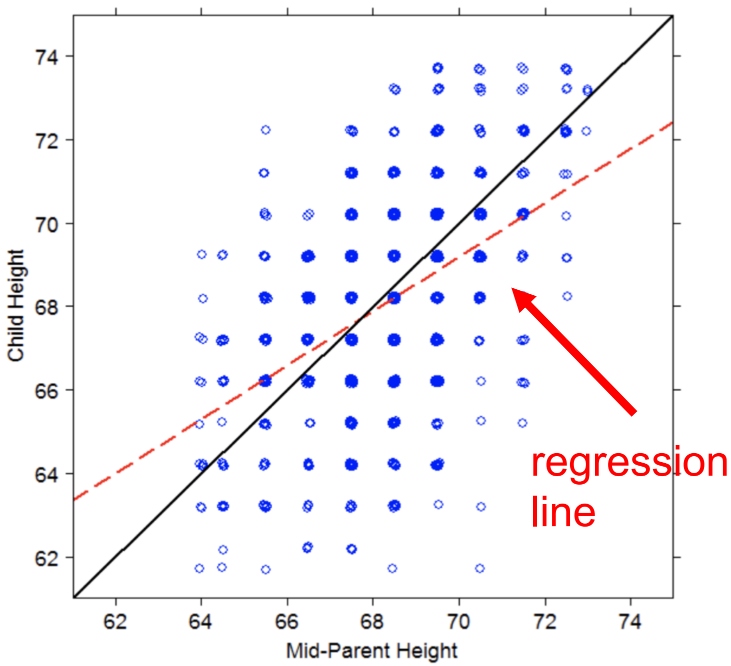
https://slideplayer.com/slide/12965344/

Arrange the density function according to the frequency distribution table for each row.

# Find a relationship that approaches (regresses to) the overall mean.

It wasn't a simple diagonal correlation.




https://slideplayer.com/slide/12965344/

## When do we use linear regression?

# When you want to understand the trend of your data

Not only biological data, but also some events in the natural world have data with quantity distributions that follow a normal distribution. Linear regression is effective when the data distribution follows a normal distribution.

# Linear Regression - Importance

**Linear regression is important because it is analytically tractable (1) and forms the basis of more sophisticated models.**
**It is important because it forms the basis for more sophisticated models.**

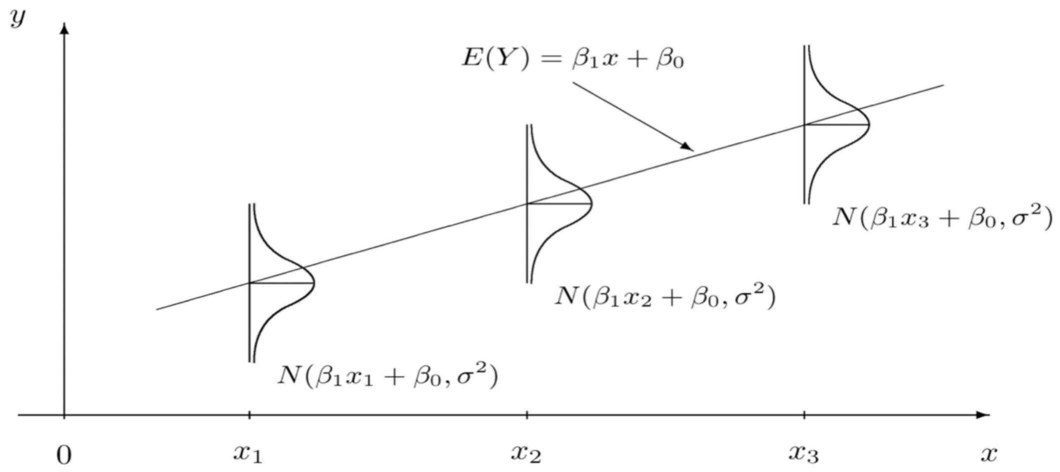
**(1) In the sense that it is easy to see what kind of change the solution will show in response to a change in a certain variable.**

# Linear Regression - Conditions

**In linear regression, the following assumptions are made.**

**(1) The explanatory variable (x) is continuous or discrete, and the objective variable (y) is continuous.**
**(2) The predicted value (y) follows a normal distribution (mean μ, standard deviation σ). (1)**

> **(1) At the i-th data point xi, the predicted value will be the mean value μi of this normal distribution.**




②のイメージ

# What kind of function is linear regression?

## When there is only one feature

Can be written in linear functions

$$\hat{y} = ax + b$$

Unknown feature

Predicted value of the correct answer label (y_pred)
In sklearn, y_pred = linear_reg.predict(X_test)
which is equivalent to

# What kind of model (function) is linear regression?

## When there is only one feature

The slope a and intercept b of this linear function are the parameters to be learned by the model.

$$\hat{y} = ax + b$$

Parameters of a linear regression model (function)

In machine learning, parameters are often expressed in terms of θ (theta) or w

**where θ0 is the intercept and θ1 is the slope (x0 is written for convenience and is equal to 1)**

$$= \theta_0 x_0 + \theta_1 x_1$$

Unknown feature

# What kind of model (function) is linear regression?

## When there is only one feature

The slope a and intercept b of this linear function are the parameters to be learned by the model.

$$\hat{y} = ax + b$$

Notice how the positions have been swapped!

In machine learning, parameters are often expressed in terms of θ (theta) or w

$$= \theta_0 x_0 + \theta_1 x_1$$

**where θ0 is the intercept and θ1 is the slope (x0 is written for convenience and is equal to 1)**

# What kind of model (function) is linear regression?

## When there are two or more features

Assuming that x1 to xn are features, the equation is as follows

$$= \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_j x_j + \ldots + \theta_n x_n \quad (x_0 = 1)$$

# What kind of model (function) is linear regression?

## When there are two or more features

There are parameters θ for [number of features + 1 intercept].

$$= \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_j x_j + \ldots + \theta_n x_n . \; (x_0 = 1)$$

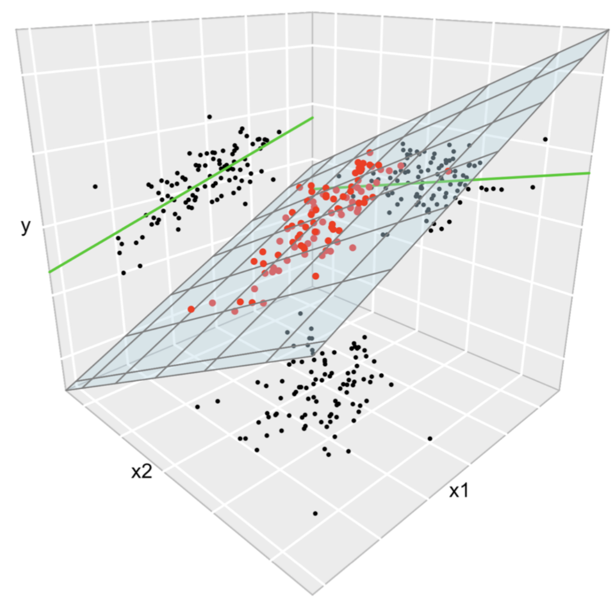
# What kind of model (function) is linear regression?

## When there is only one feature

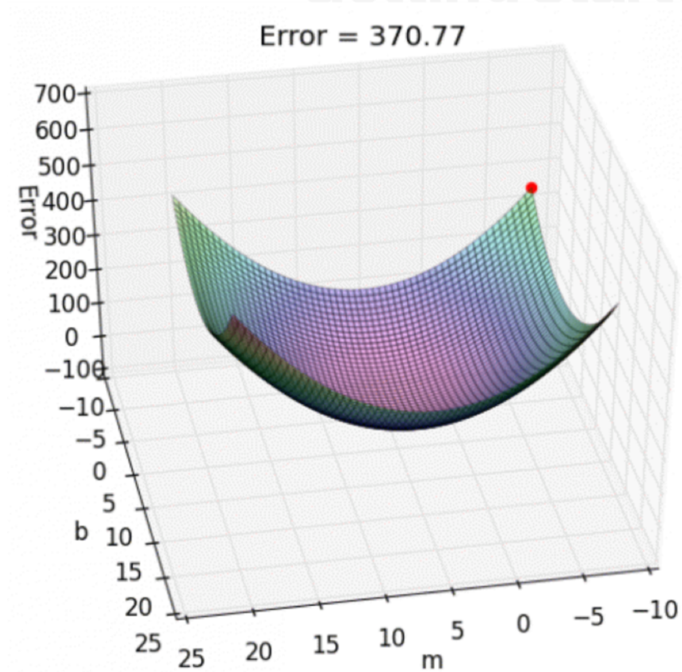
When there is only one feature, it is called a single regression analysis.
Fit a straight line of linear regression to the data.

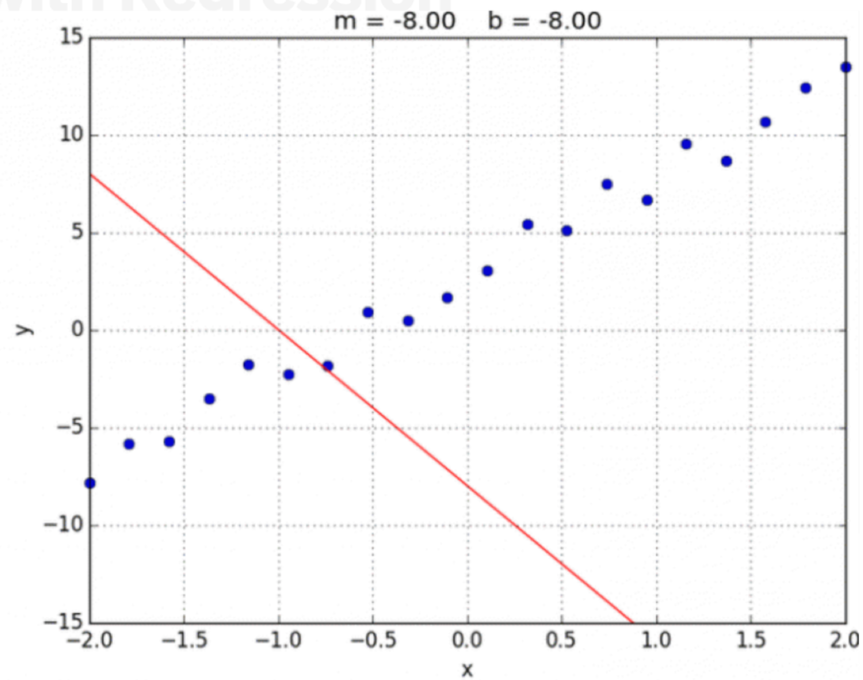
## When there are two or more features

When there are more than two features,
 it is called multiple regression analysis.
If there are two features,
we fit a linear regression plane to the dat
If there are three or more features,
it is difficult to visualize.

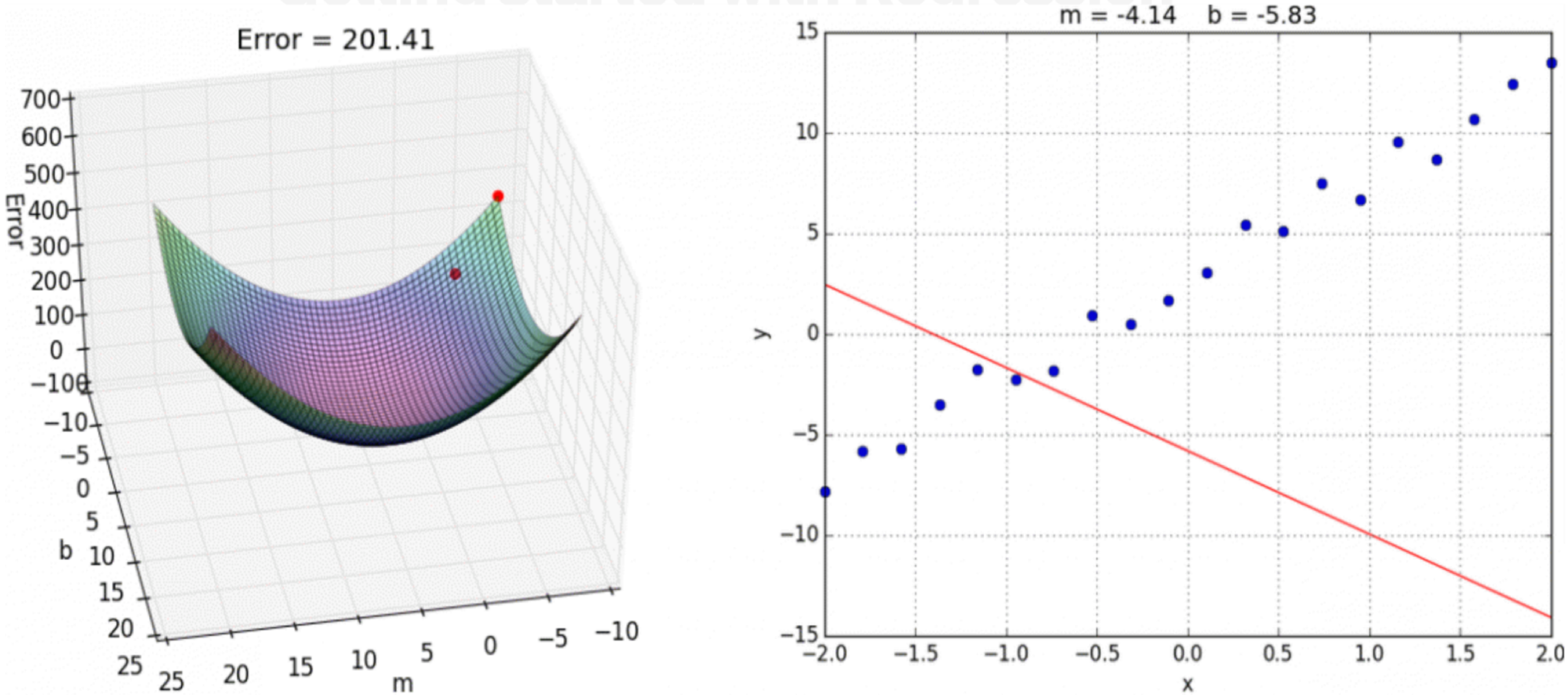


# Understand the feelings of the relationship between these two graphs.

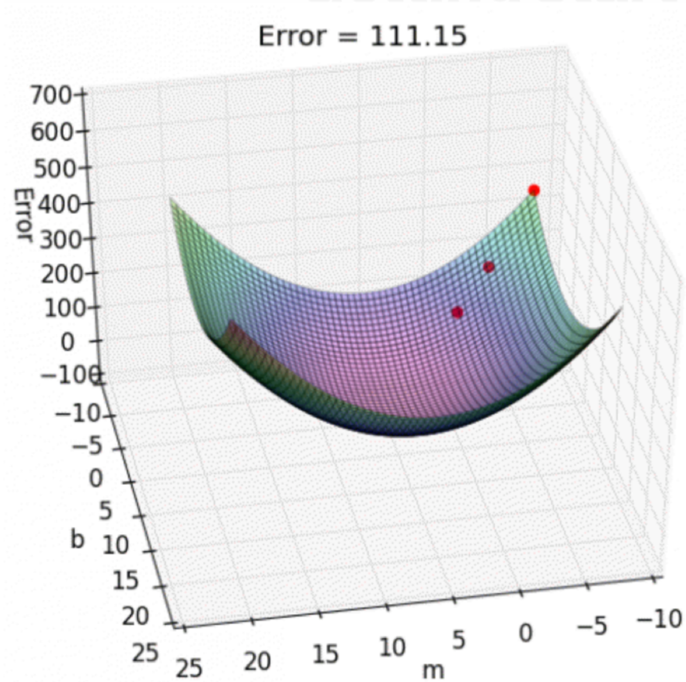


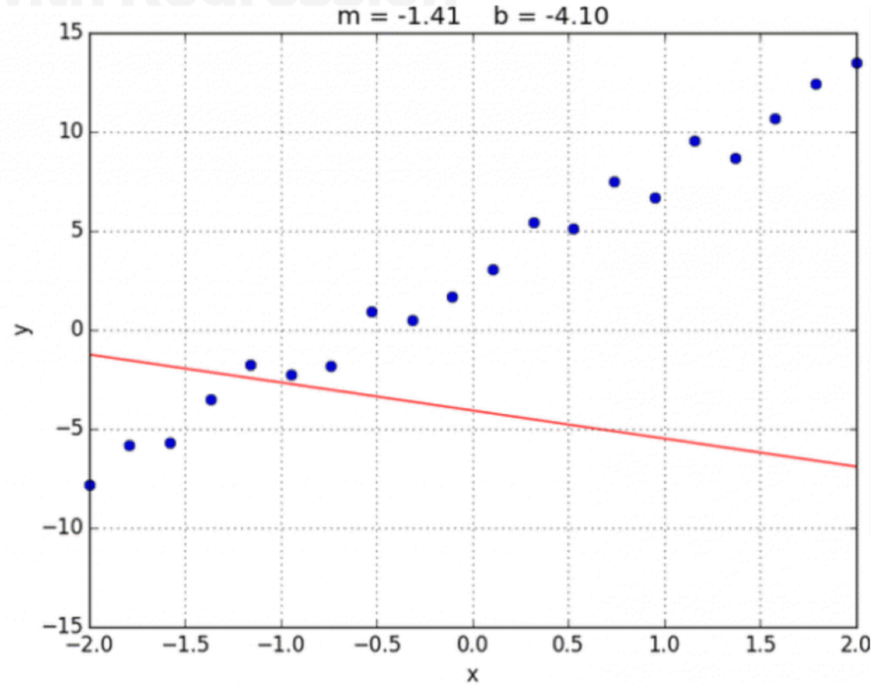



# Understand the feelings of the relationship between these two graphs.

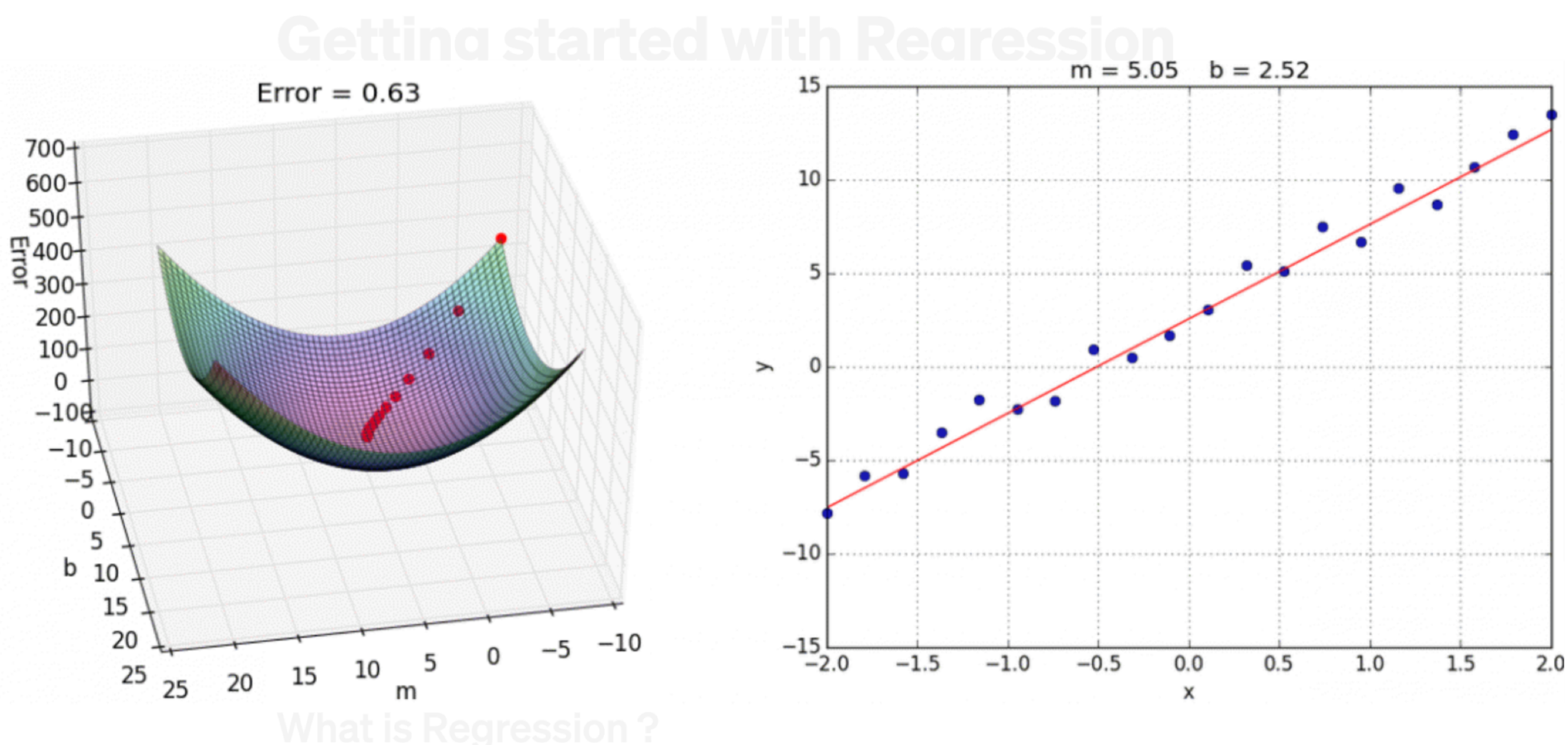


# Understand the feelings of the relationship between these two graphs.

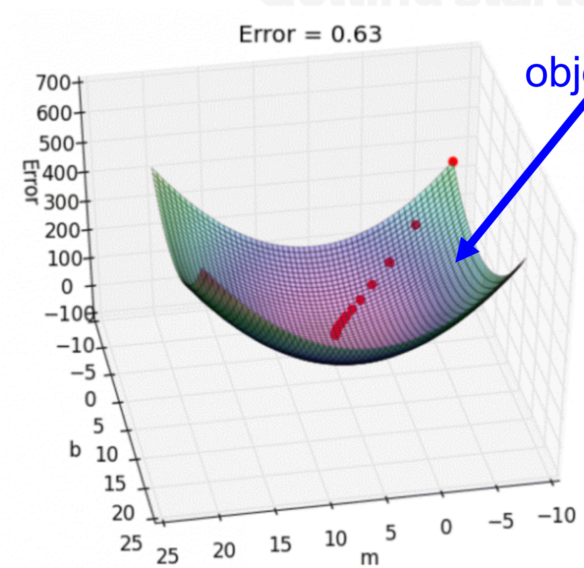


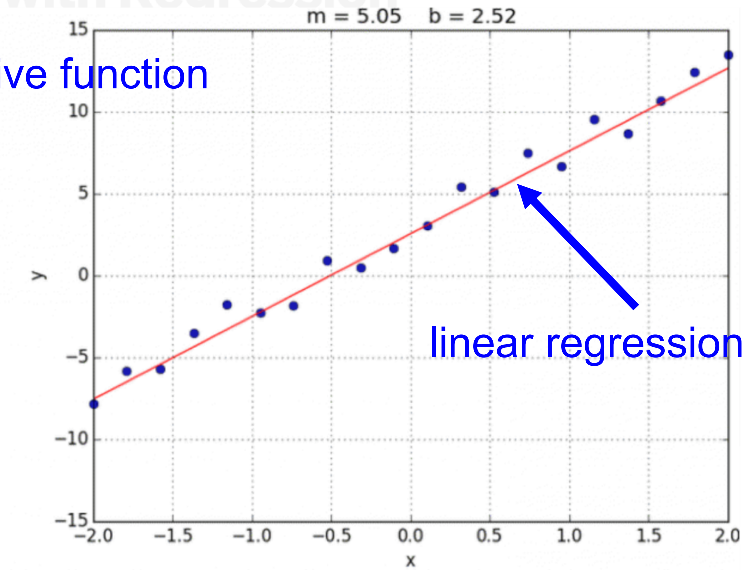



# Understand the feelings of the relationship between these two graphs.



# What is this graph?



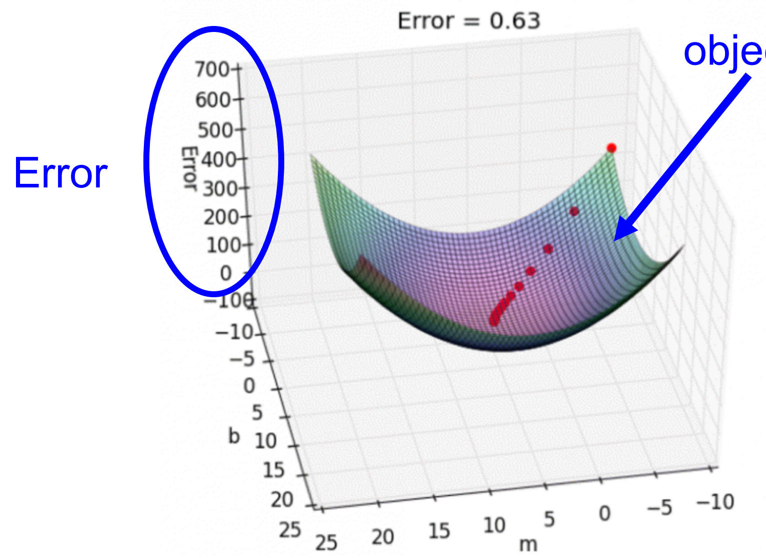



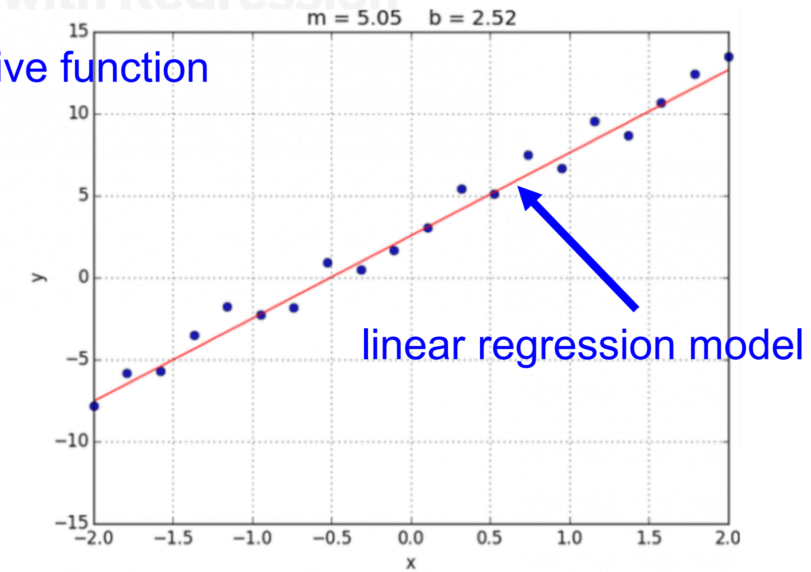

Objective Function
Vertical axis:b Horizontal axis:m Height:Error

Linear regression model (function)
m:Slope b:Intercept

## How do I look at these graphs?

# How do I look at the graph?




Objective Function
Vertical axis:b Horizontal axis:m Height:Error

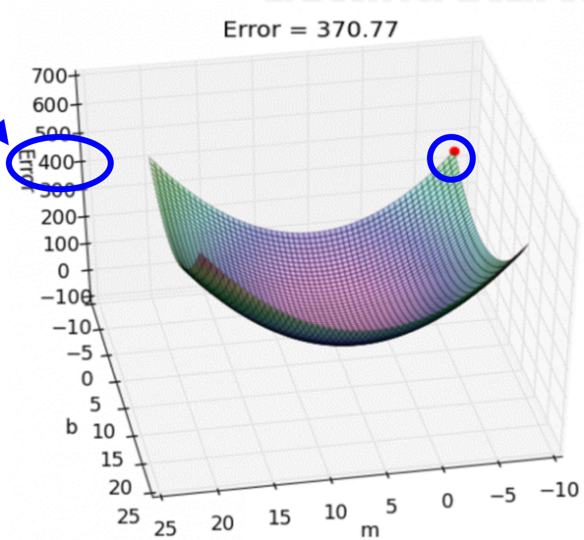
Linear regression model (function)
m:Slope b:Intercept

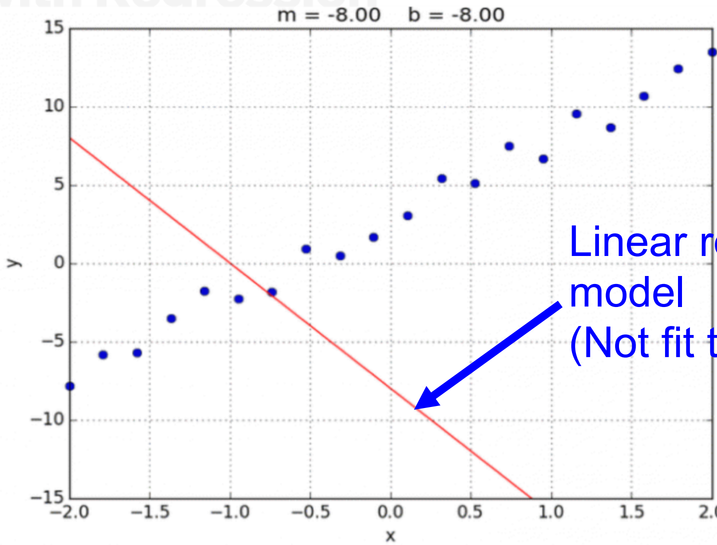
Whenever the value of Error in the objective function changes, the position of the linear regression model changes.

https://medium.com/@savannahar68/getting-started-with-regression-a39aca03b75f

# How do I look at the graph?

Error value



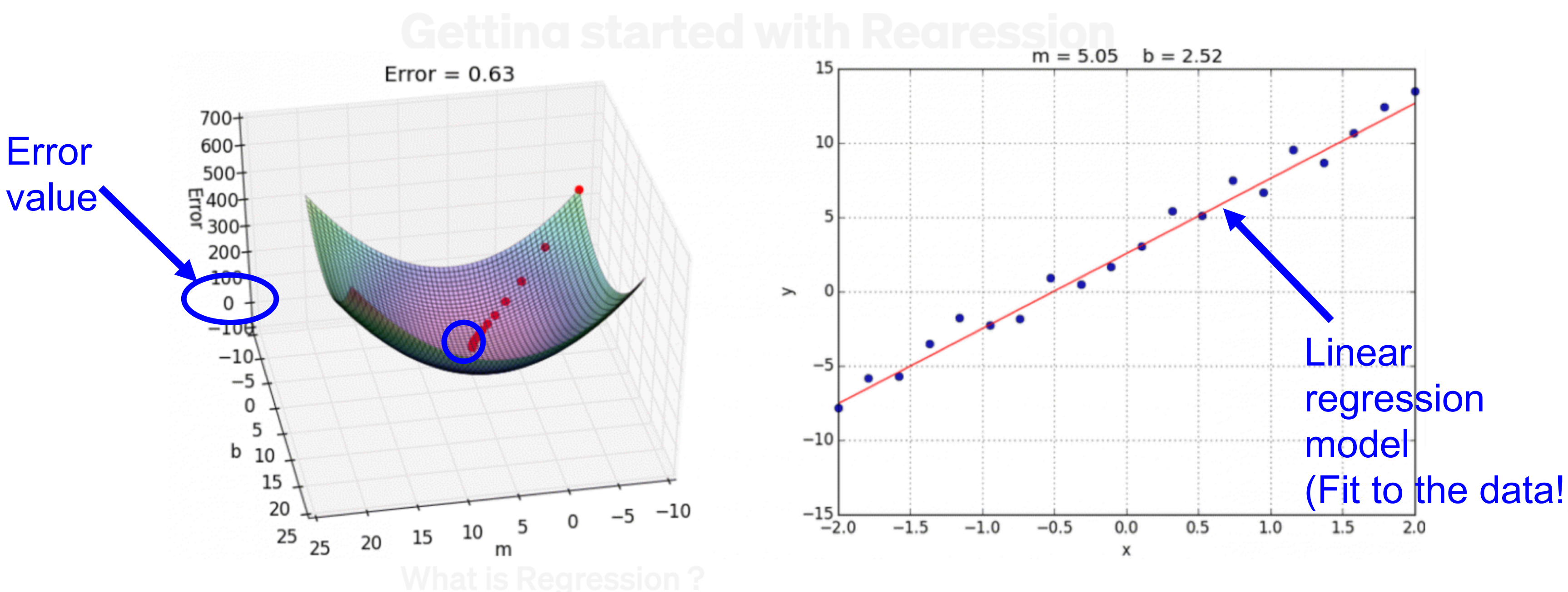

Linear regression model (Not fit to data)

Objective Function
Vertical axis:b Horizontal axis:m Height:Error

Linear regression model (function)
m:Slope b:Intercept

Optimizing the value of Error in the objective function
A good linear regression model (slope and intercept) can be obtained.

# How do I look at the graph?




Objective Function
Vertical axis:b Horizontal axis:m Height:Error

Linear regression model (function)
m:Slope b:Intercept

Optimizing the value of Error in the objective function

A good linear regression model (slope and intercept) can be obtained.

# Linear Regression – The Flow

## The Flow of the Linear Regression

(1) Formulate an equation (hypothetical function) to derive the predicted value

(2) Calculate the error between the target variable and the predicted value.

(3) Set up a problem to minimize this error and formulate an equation (objective function).

(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).

(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.

# Linear Regression – The Flow

**HousePrice data (week4_work2)**
**Suppose we have a HousePrice data set.**
**Let's choose one explanatory variable x (GrLivArea) for the objective variable y (SalePrice) and plot the relationship between the two variables.**

**A linear relationship is likely to exist.**
**Looking at the trend of the plotted data, it seems that a straight line can be drawn.**

# Linear Regression – The Flow

**When there is a linear relationship between variables x and y, it can be expressed in this way.**

$$\hat{y}_i = \beta_0 + \beta_1 x_i \qquad i = 1, 2, \ldots, n$$

**Let's use this equation to draw a straight line.**
**The parameter β1 controls the slope of the straight line.**
**In other words, it shows how much change per unit of variable xi corresponds to a change in variable yi.**
**The parameter β0 is the intercept that shows the point of intersection with the y-axis when xi = 0.**

# Linear Regression – The Flow

## The Flow of the Linear Regression

(1) Formulate an equation (hypothetical function) to derive the predicted value

(2) Calculate the error between the target variable and the predicted value.

(3) Set up a problem to minimize this error and formulate an equation (objective function).

(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).

(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.

# Linear Regression – The Flow

**If we have a straight line that fits well to the HousePrice dataset (train_X) that we have, we may be able to predict the corresponding y (i.e. the unknown SalePrice) when we get the unknown dataset (test_X).**

**Is the straight line a good fit?**
**There is a large error between the prediction yi (100000) and the target variable yi (200000) in the sample xi (3450) with the explanatory variables.**

# Linear Regression – The Flow

**We want to evaluate this error and minimize it.**
**To evaluate how much error there is overall, we add up all the errors.**
**The smaller the sum of these errors, the better the line fits.**

# Linear Regression – The Flow

**The formula for evaluating the error can be written as follows (this method of evaluation is called the least squares method).**

$$\mathcal{L}(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad i = 1, 2, \ldots, n$$

**Since each point is scattered above and below the line, the square cancels out the positive and negative.**

**DIVER is formulated as a multiple regression model, but this is a single regression model.**

# Linear Regression – The Flow

## The Flow of the Linear Regression

(1) Formulate an equation (hypothetical function) to derive the predicted value

(2) Calculate the error between the target variable and the predicted value.

(3) Set up a problem to minimize this error and formulate an equation (objective function).

(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).

(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.

# Linear Regression – The Flow

## The evaluation index is the mean squared error (MSE).

**When we take the average of all the errors, the least-squares method is replaced by the mean-squared error (= variance).**
**We will use this mean squared error as an evaluation measure for linear regression.**

**Furthermore, this evaluation index divided by two is the Cost Function of linear regression. The equation of the straight line used to calculate the predicted value is called the hypothesis function.**
**The objective function is divided by two for convenience of calculation, since it will be subject to partial differentiation later.**

線型結合

Hypothesis: $h_\theta(x) = \boxed{\theta_0 + \theta_1 x}$

特徴量方向の足し算

DIVERでは
$\theta_0 x_0 + \theta_1 x_1 \cdot$
$(x_0 = 1)$

Parameters: $\theta_0, \theta_1$

サンプル(m個)方向の足し算

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \boxed{\sum_{i=1}^{m}} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

# Linear Regression – The Flow

## The goal is to minimize the value of $J(\theta_0, \theta_1)$

**A problem that seeks the minimum value of an objective function is generally called an optimization problem (a problem to minimize or maximize some function under given conditions).**

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: $\theta_0, \theta_1$

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

J(Θ₀,Θ₁)を最小化するようなパラメータΘ₀,Θ₁を求める

# Linear Regression – The Flow

## The Flow of the Linear Regression

**(1) Formulate an equation (hypothetical function) to derive the predicted value**

**(2) Calculate the error between the target variable and the predicted value.**

**(3) Set up a problem to minimize this error and formulate an equation (objective function).**

**(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).**

**(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.**

# Linear Regression – The Flow

**Where does the optimal solution lie in the function?**

**Since this is a problem of finding the minimum value, the optimal solution is the point with the smallest function value in the graph.**

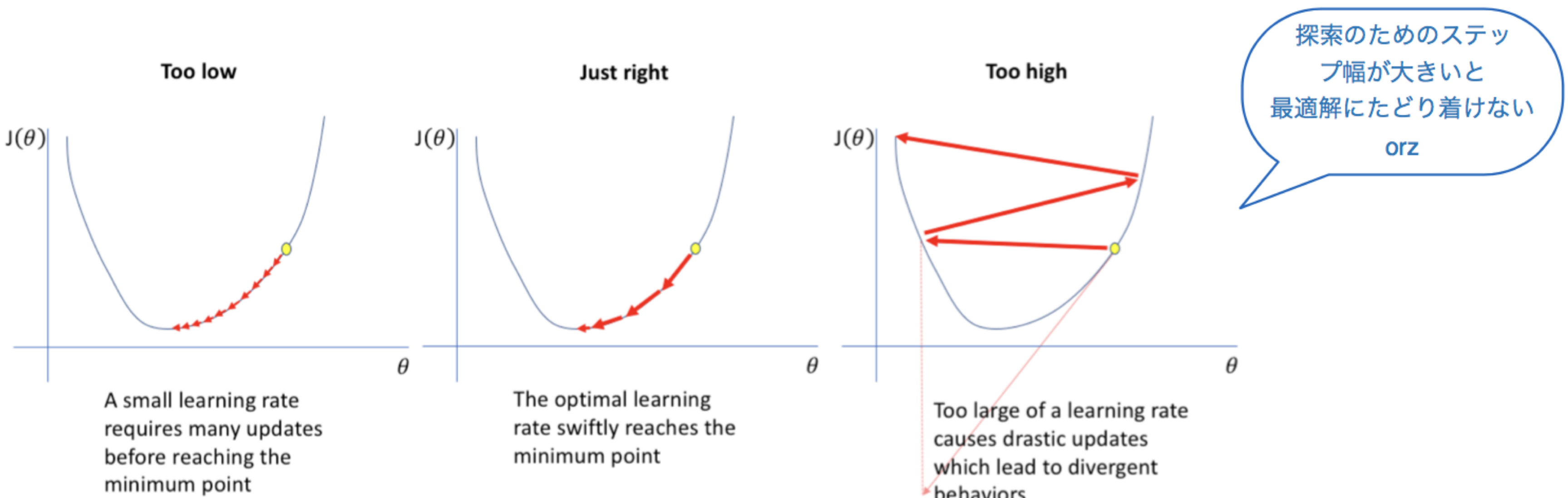
**A feature of the steepest descent method (1) as an "optimization method" is that it sums up all the errors in the training data and updates the parameters. This is the simplest and most classical of the several types of gradient descent methods.**

**(1) Distinguish between the steepest descent method as an "analytical method".**
https://en.wikipedia.org/wiki/Method_of_steepest_descent#Extensions_and_generalizations

# Linear Regression – The Flow

## Finding the optimal solution

**Since the objective function J(θi) is calculated for each training session**
**Since the objective function J(θi) is calculated for each training session, it is not possible to know in advance what shape the graph will take. In the optimization problem here, the**
**Minimum value of the function ( J(θi) = 0 ): Search for a point where the slope of the tangent line becomes 0.**
**is obtained by search.**
**(In the case of a constrained minimization problem, the minimum value within the constraints becomes the optimal solution.)**

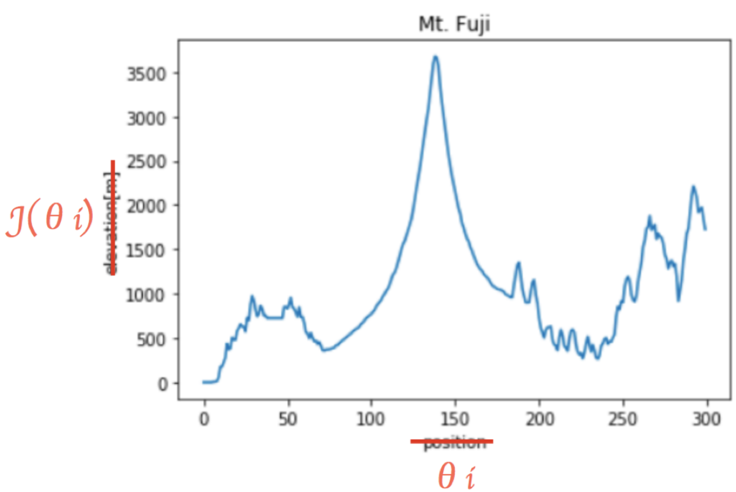



https://www.jeremyjordan.me/nn-learning-rate/

# Linear Regression – The Flow

## What is the Fastest Descent Method?

**A type of iterative algorithm that starts with a suitable initial point and updates the point as follows.**




※ week2 授業課題2 富士下山問題 より

$$\theta_{i+1} := \theta_i - \alpha \frac{\partial J(\theta_i)}{\partial \theta_i}$$

$$\frac{\partial J(\theta_i)}{\partial \theta_i}$$

**denotes the search direction, and we proceed in this direction.**

**We expect the i + 1 iteration point to be closer to the solution than the i-th iteration point. α is a scalar, a step width that controls how far to go in the search direction, also known as the learning rate. As the name implies, the steepest descent method searches for a solution using the direction (gradient) that is the smallest.**

# Linear Regression – The Flow

## The Flow of the Linear Regression

**(1) Formulate an equation (hypothetical function) to derive the predicted value**

**(2) Calculate the error between the target variable and the predicted value.**

**(3) Set up a problem to minimize this error and formulate an equation (objective function).**

**(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).**

**(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.**

# Linear Regression – The Flow

## Once the assumed function has the optimal parameters

**Let's try to estimate it using the assumed function with the variable θ last updated.**
**The equation looks the same, but the value of θ is no longer unknown.**

np.random.randで一様分布から
サンプリングした乱数（[0,1]の範囲）

**学習前** $$h_\theta(x) = \boxed{\theta_0} + \boxed{\theta_1} x$$ ← 誤差の算出に使用する
（x：学習データ）

**学習後** $$h_\theta(x) = \boxed{\theta_0} + \boxed{\theta_1} x$$ ← 推定に使用する
（x：テストデータ）

最適化された値

# Linear Regression – The Flow

## The Flow of the Linear Regression

(1) Formulate an equation (hypothetical function) to derive the predicted value

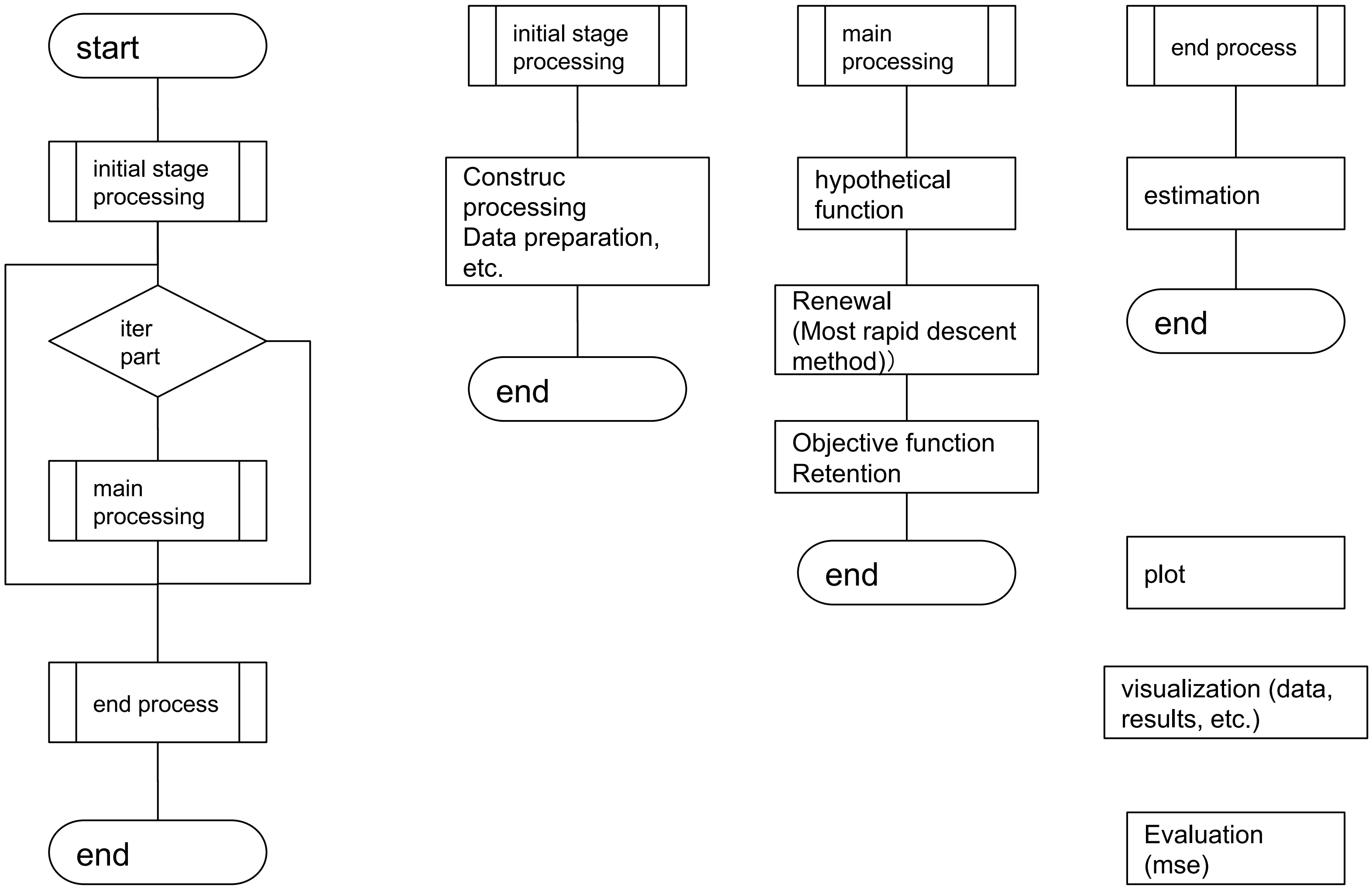
(2) Calculate the error between the target variable and the predicted value.

(3) Set up a problem to minimize this error and formulate an equation (objective function).

(4)Find the optimal solution of the objective function in an exploratory manner (steepest descent method).

(5)When the optimal solution is reached, the optimal parameters of the assumed function are obtained.

# The Flow



start
initial stage processing
iter part
main processing
end process
end
initial stage processing
Construc processing Data preparation, etc.
end
main processing
hypothetical function
Renewal (Most rapid descent method))
Objective function Retention
end
end process
estimation
end
plot
visualization (data, results, etc.)
Evaluation (mse)

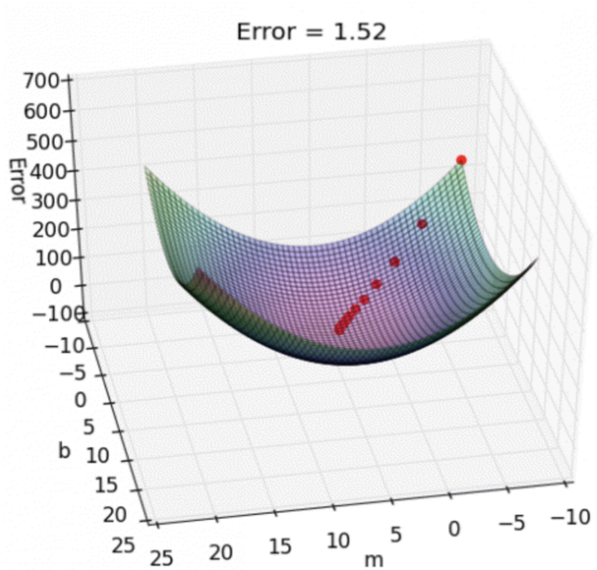
# Optimization

When we want to solve a real-world problem using mathematics, we need to formulate it. The formulated problem is called an optimization problem.

# What is optimization?

**An optimization problem is a problem of finding a solution that minimizes (maximizes) the value of an objective function under a given set of constraints.**

Minimizing or maximizing the value of the objective function is called optimization.

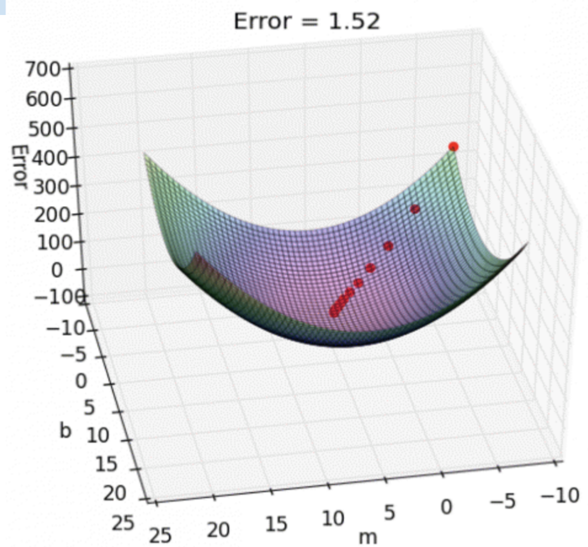


# What is optimization?

## The objective function this time

Mean Squared Error (MSE))

$$J = \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2y_i(ax_i + b) + (ax_i + b)^2)$$




$$J_i = (y_i - \hat{y}_i)^2 \qquad = (y_i - (ax_i + b))^2$$

$$\hat{y} = ax + b$$

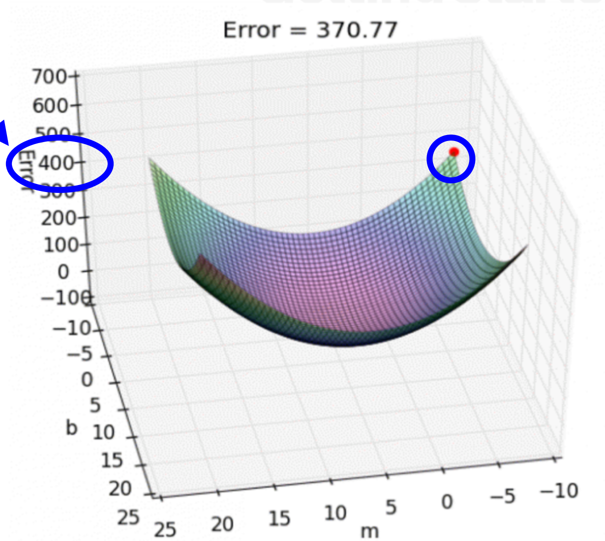
Objective Function

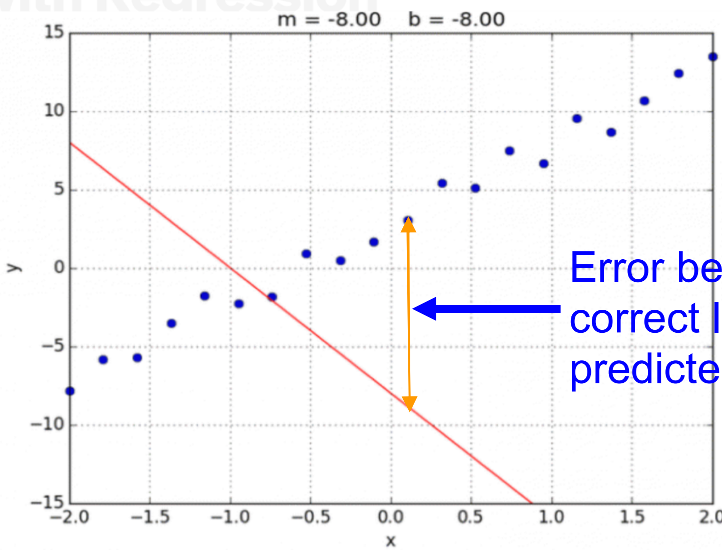
Vertical axis:b Horizontal axis:m Height:Error

The objective function this time represents the square of the error between the correct label y and the predicted value y_hat

# What is optimization?

Error value



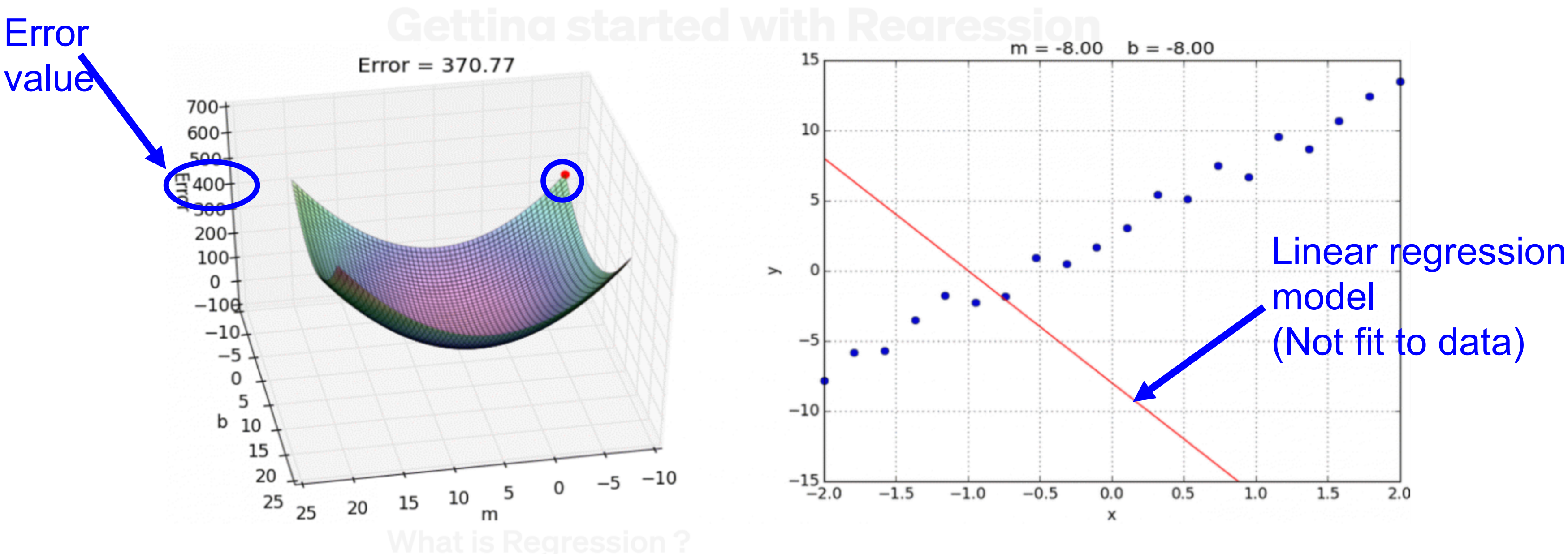

Error between correct label and predicted value

Objective Function
Vertical axis:b Horizontal axis:m Height:Error

Linear regression model (function)
m:Slope b:Intercept

The objective function this time is the square of the error between the correct label y and the predicted value y_hat which represents the square of the error between

# What is optimization?




Objective Function
Vertical axis:b Horizontal axis:m Height:Error

Linear regression model (function)
m:Slope b:Intercept

Optimizing the value of Error in the objective function

A good linear regression model (slope and intercept) can be obtained.

# Is it useful in practice?

事例

- An internship exam at a Fintech company asked me to predict the price of financial products using multiple regression analysis

- Using multiple regression analysis to measure the effectiveness of web marketing measures at a certain z***.

- Using multiple regression analysis to predict the number of monthly word-of-mouth posts at Retty

https://www.nomura.co.jp/terms/japan/ka/A01941.html

# Frequently Asked Questions

Q. Linear regression models are less expressive than neural networks because of their simple structure, but do they have any use in practice?

A. Because of the simple structure of the model, in multiple regression analysis, we can see which features of the input are working for prediction and how much, so we can make decisions based on that.

# Frequently Asked Questions

Identify which features have a significant impact.

↓

We can make decisions to adjust its features.

The adjustment of these features is more important than the prediction accuracy of the model.

https://engineer.retty.me/entry/2019/12/17/120000

# Points to note when using linear regression

## Multicollinearity

In regression analysis, if the correlation between features is high, the parameter θ cannot be obtained correctly (the variation becomes large). Therefore, it is generally necessary to check the correlations among the features and exclude the features with high correlations before performing regression analysis, but as the number of features increases, the time and effort required to do so increases dramatically.

https://webbeginner.hatenablog.com/entry/2016/05/29/072806

https://socinuit.hatenablog.com/entry/2020/06/26/013753

## Deal with

- Variable increase/decrease method (stepwise method)
- Perform multiple regression analysis by excluding features one by one, and look for combinations that can be resolved.

# Linear Regression of scikit-learn

**Let’s first have a look at the one used until now with the help of the scikit-learn library.**

[Scikit-learn’s LinearRegression Class]

Let’s use Google Colab. and check it quickly

# Sprint 3 – Scratch Linear Regression

**Explanation about this Sprint is given but please try it on your own first.**

[Sprint 3 – Scratch Linear Regression]

Please work on your own after class and submit your assignments on DIVER.

# Sprint 3 – Sample Code

**A Sample Code of this Sprint is given but please try it on your own.**

[Sprint 3 – Scratch Linear Regression]

Please work on your own after class and submit your assignments on DIVER.

# ToDo by next class

Next class will be held on Zoom : Thursday 20 May 2021

🚨ToDo: Sprint 4 – Scratch Logistic Regression
https://diver.diveintocode.jp/curriculums/1645

# Check-out

**3 minutes** Please post the following point to Zoom chat.

**Q. Current feelings and reflections**
(joy, anger, sorrow, anticipation, nervousness, etc.)

# Thank You For Your Attention