Week_2

Summer Workshop 2023
Hallym University Academic Club

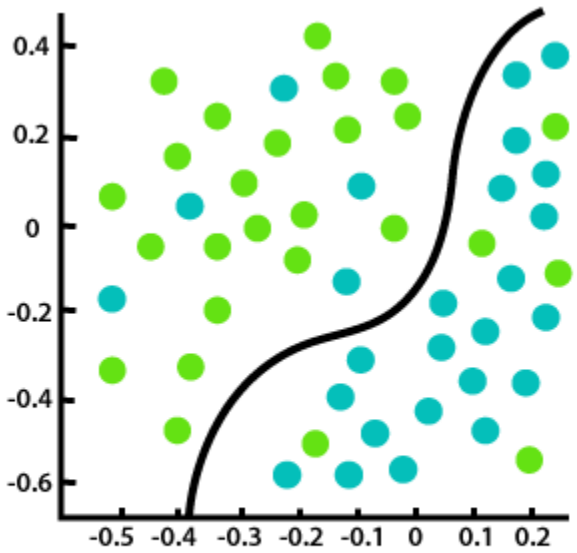
Hallym University Department of Big Data in the school of Software
Tae-Hoon Her

HALLYM UNIVERSITY

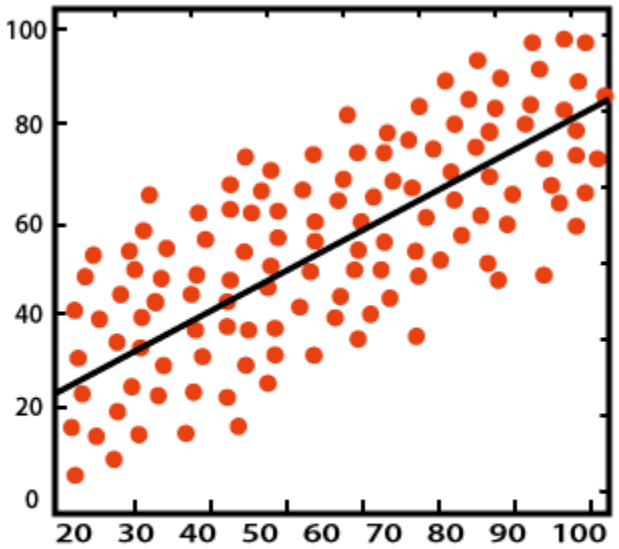
This course

- Concept of regression
- Gradient descent
- Implementing regression in Python

Regression & Classification



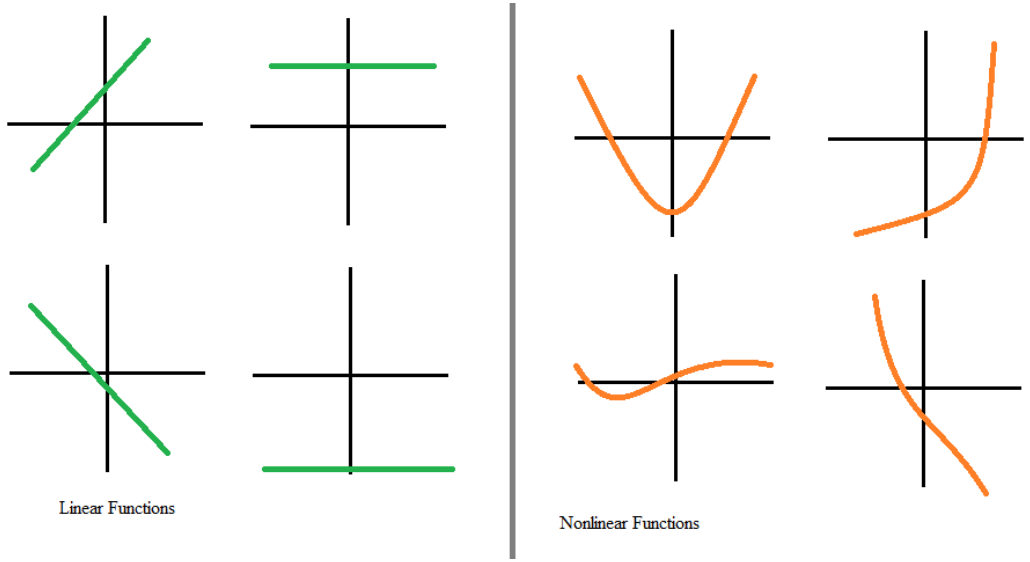
Classification



Regression

Linear Regression

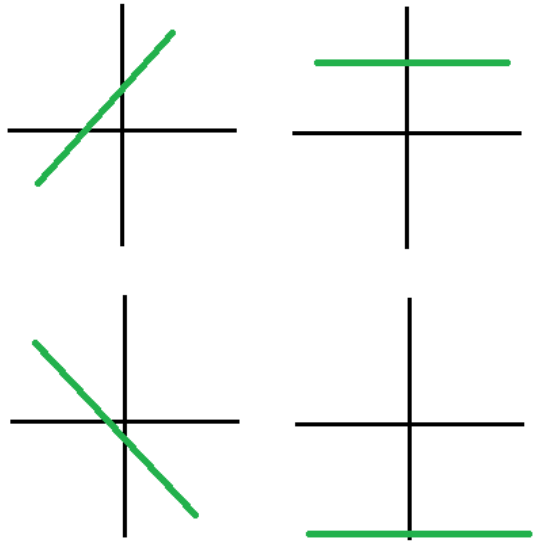
- Regression can generally be defined as the problem of finding the best-fitting line or curve that explains a set of data points plotted in a two-dimensional space
- In the context of y = f(x), where the output y is a real number and the input x is also a real number, predicting the function f(x) is considered regression



Linear Regression

- Models the linear relationship between input variables and output variables
- Based on the assumption that each change in the input variables has a constant effect on the output variables
- As the size of the house increases, the price tends to increase linearly

$$y=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta nxn+\epsilon$$

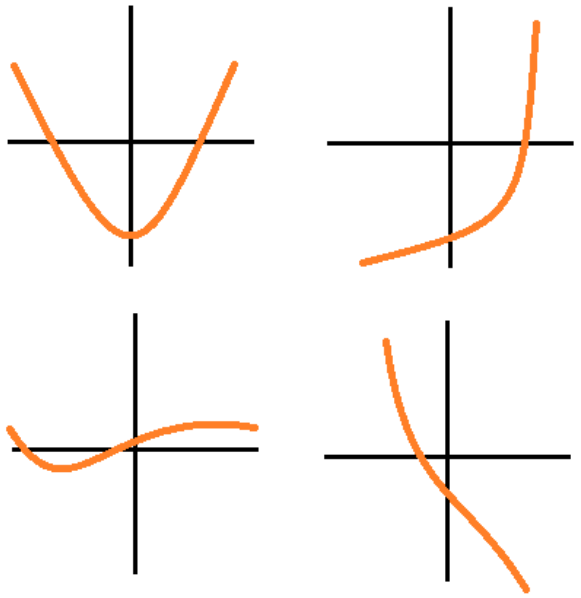


Linear Functions

NonLinear Regression

- Models the nonlinear relationship between input variables and output variables

- Used when the changes in the output variable do not proportionately correspond to the changes in the input variable



Nonlinear Functions

Example of linear regression

1. Investigating the relationship between parents' height and children's height
2. Predicting housing prices based on the area of the property
3. Forecasting unemployment rates based on age
4. Analyzing the relationship between study hours and academic grades
5. Predicting program execution time based on CPU speed

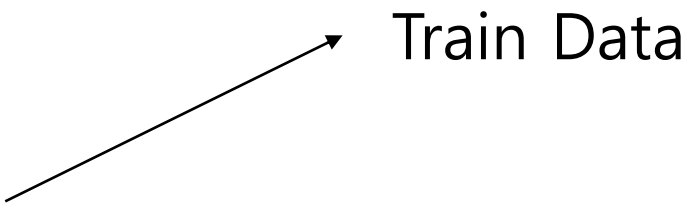
Introduction to Linear Regression

- Equation of a straight line: $f(x) = mx + b$
- Linear regression is the problem of finding the best-fitting slope (weight) and y-intercept (bias) that accurately describe the input data

- The basic form of linear regression is represented as $f(x) = Wx + b$
  - The slope is referred to as the weight
  - The y-intercept is referred to as the bias

Example of Linear Regression

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75



What is the height when the weight is 225(lbs)?

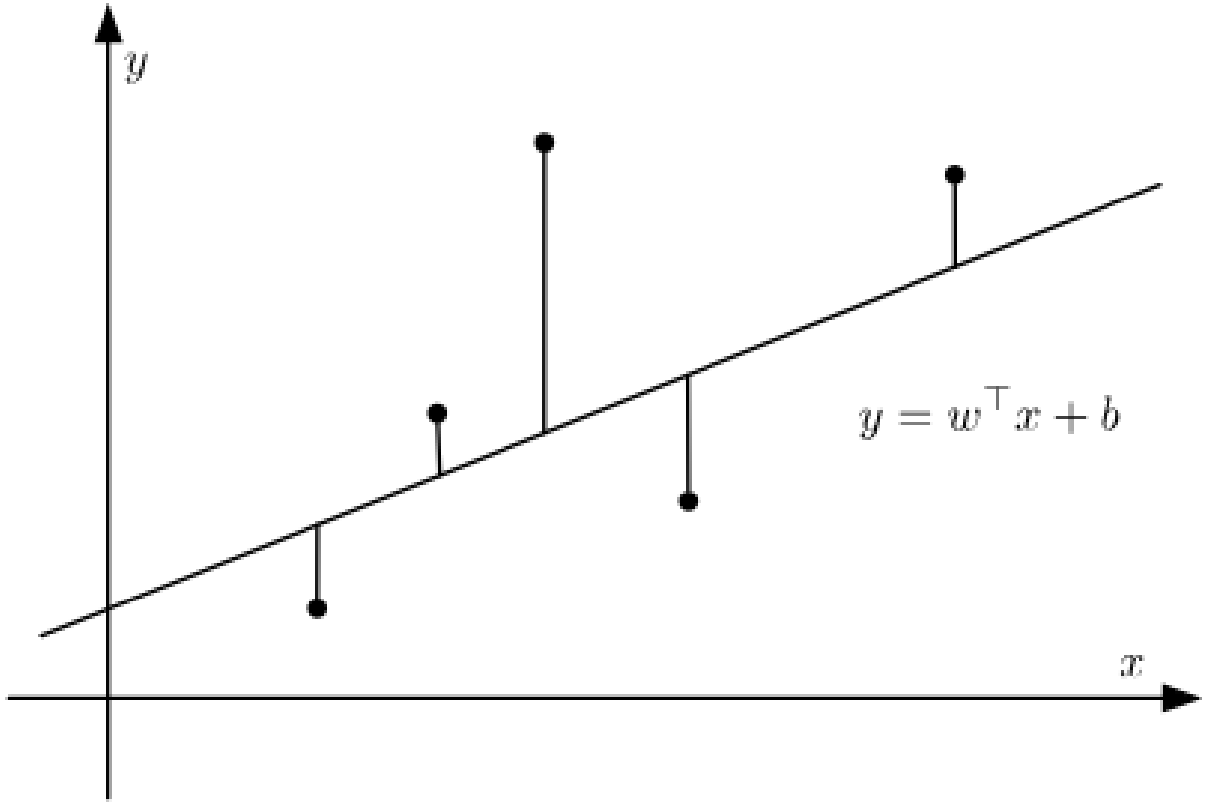
Types of Linear Regression

- Simple Linear Regression: Simple linear regression is a type of linear regression where there is only one independent variable (x)

$$f\ x = wx + b$$

- Multiple Linear Regression: Multiple linear regression is a type of linear regression where there are multiple independent variables

Loss Function



- On the graph, each point represents an observation
- The drawn line represents the linear regression model
- The drawn line predicts the output value for a given input value
- The distance between the line and the point represents the difference, or error, between the model's predicted value and the actual value

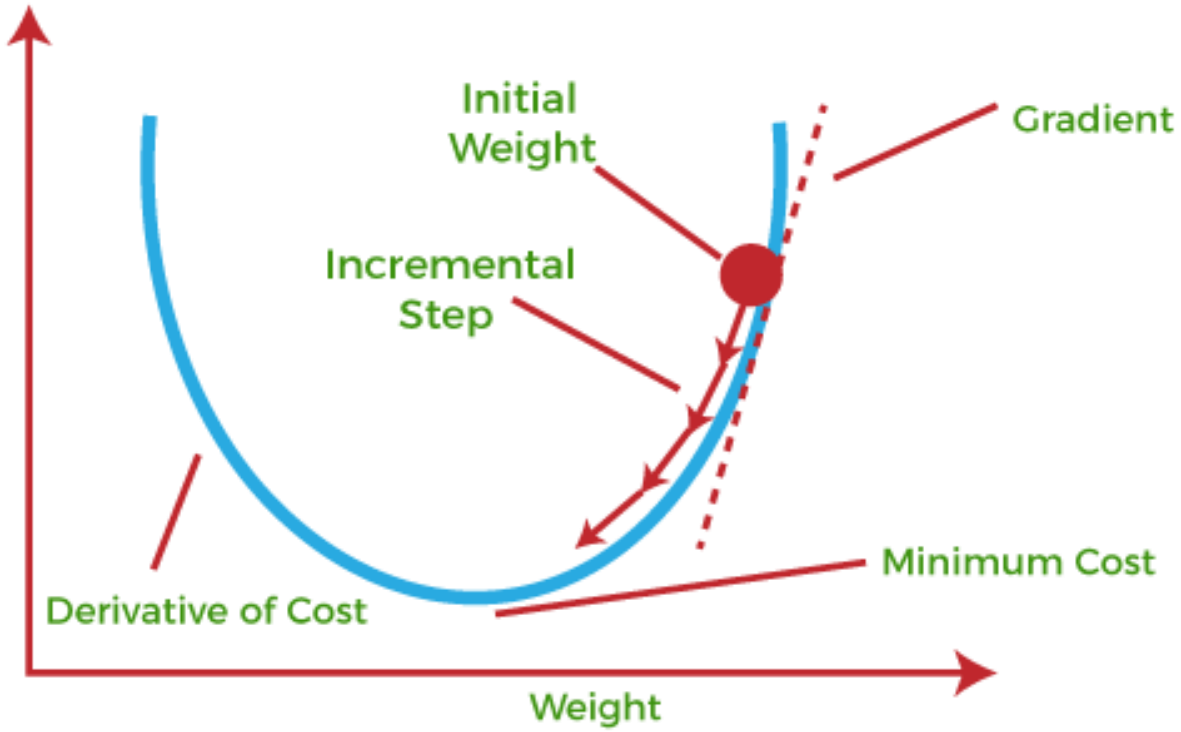
Loss Function

The sum of the squared distances between the line and the data is called the loss function or cost function

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

- The most commonly used loss function in linear regression is the Mean Squared Error
- 'y' is the actual value and 'y hat' is the value predicted by the model
- The error is calculated by squaring the difference between the actual and predicted values and averaging this over all data points

Gradient Descent



- The x-axis represents the model's parameters and the y-axis represents the loss function
- The goal is to find the minimum value of the loss function
- The model's parameters start with an arbitrary value
- The derivative, or gradient, of the graph is calculated