Introduction to Supervised Learning: Concepts and Applications

Gusti Triandi Winata

Community Researcher of RL & LLM
Cohere for AI

bangkit

About Me

Gusti Triandi Winata

Latest Work Experiences:

- Researcher, Cohere for AI **2024 - present**
- MLE Consultant, Freeport Indonesia **2023 - 2024**
- Mid. MLE, eFishery **2022 - 2023**
- OSS Fellowship with Adobe, Major League Hacking **2021**

Education:

- Bandung Institute of Technology Graduated at **2021**
  Bachelor of Electrical and Computer Engineering
  Was a Bangkit Graduate of the first batch (2020)!

bangkit

Ground Rules

Observe the following rules to ensure a supportive, inclusive, and engaging classes

Give full attention in class

Mute your microphone when you're not talking

Keep your camera on

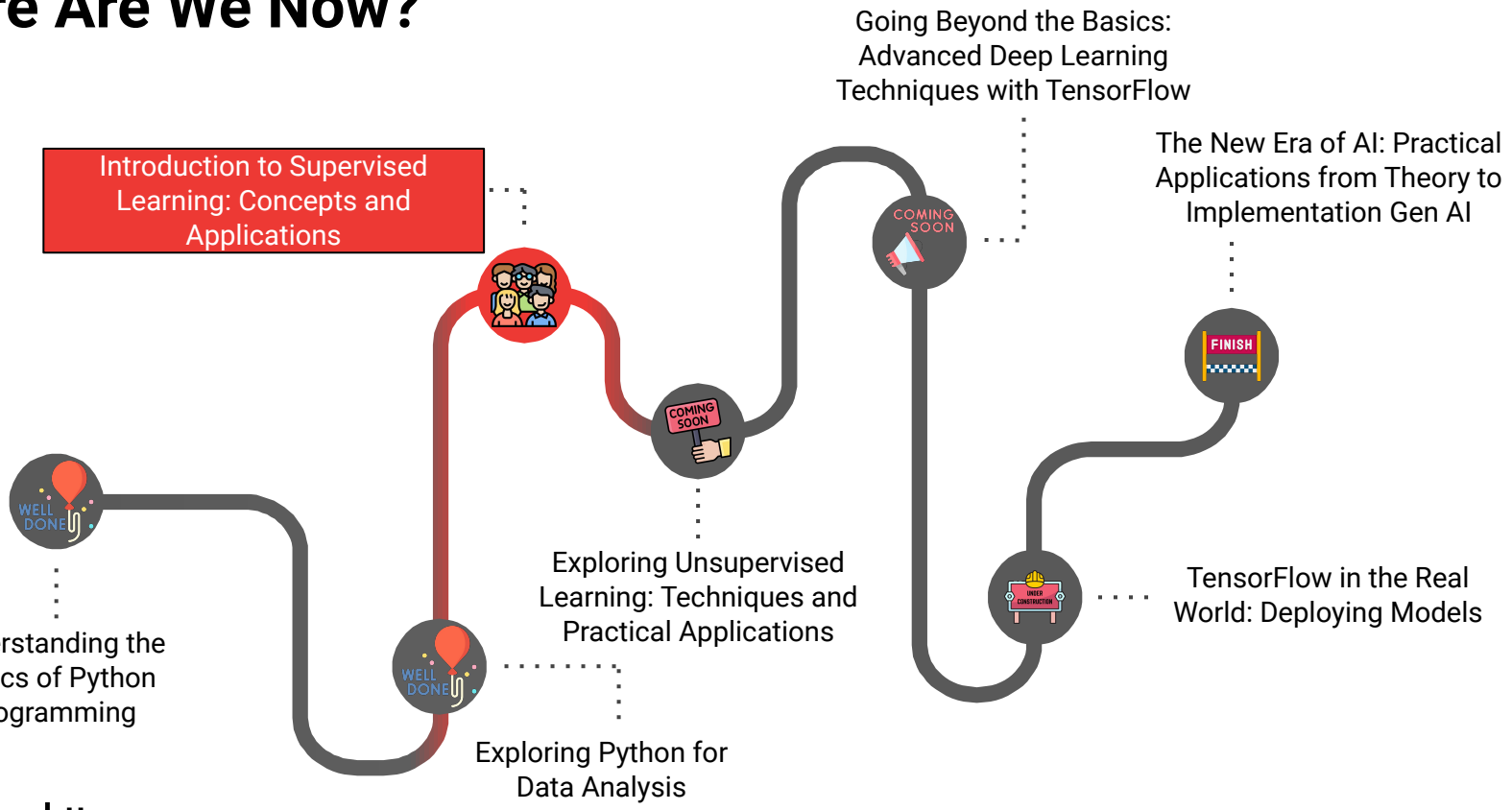
Turn on the CC Feature on Meet

Use raise hand or chat to ask questions

Make this room a safe place to learn and share

bangkit

Where Are We Now?

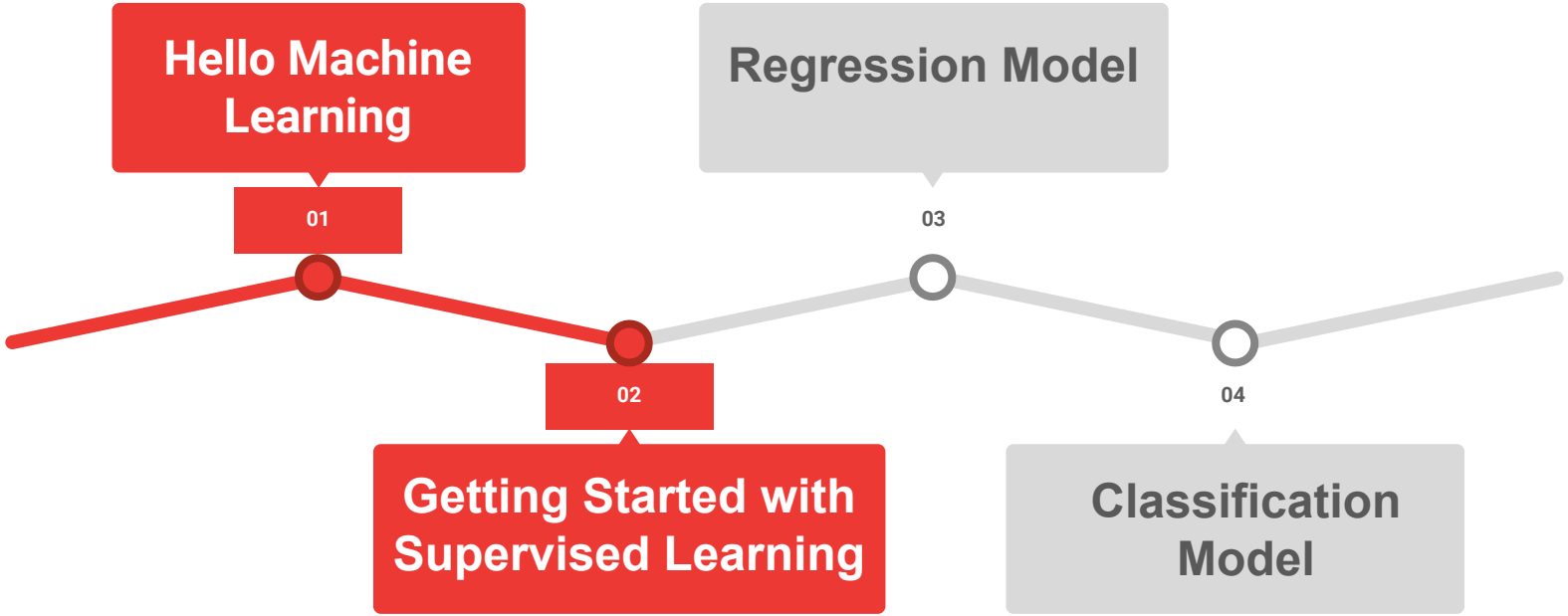


bangkit

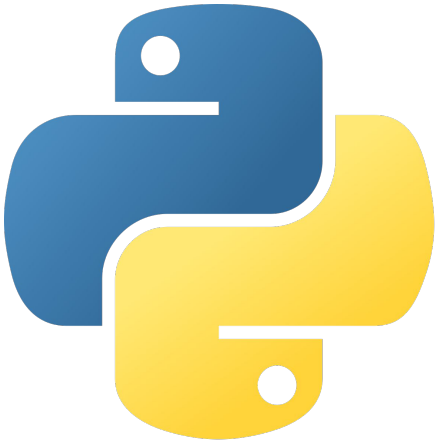
Learning Objectives

- Identify basic concepts in machine learning
- Conceptualising supervised learning
- Comparing algorithm in supervised learning

Today's Agenda



bangkit



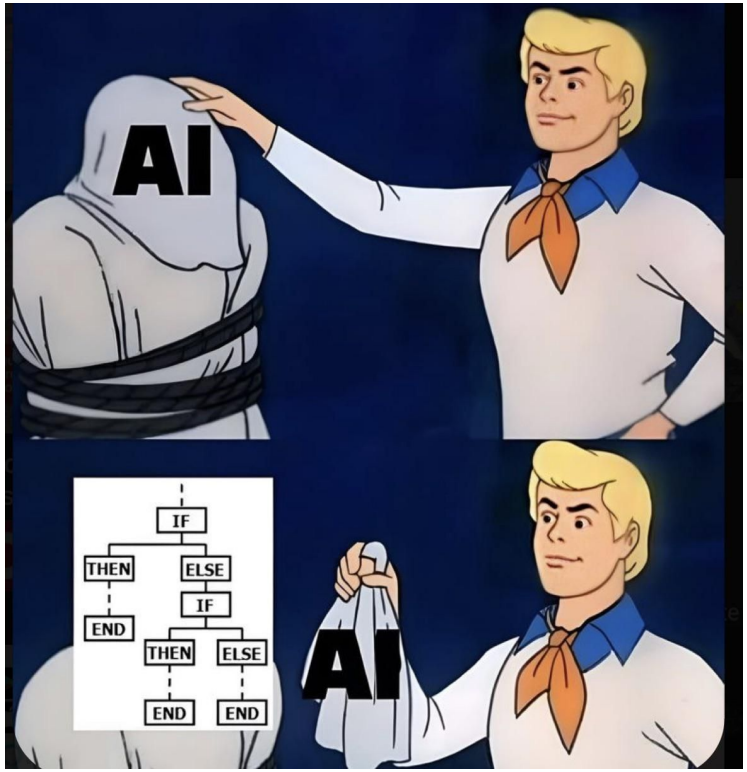
Condition Evaluations

Condition of Comparison

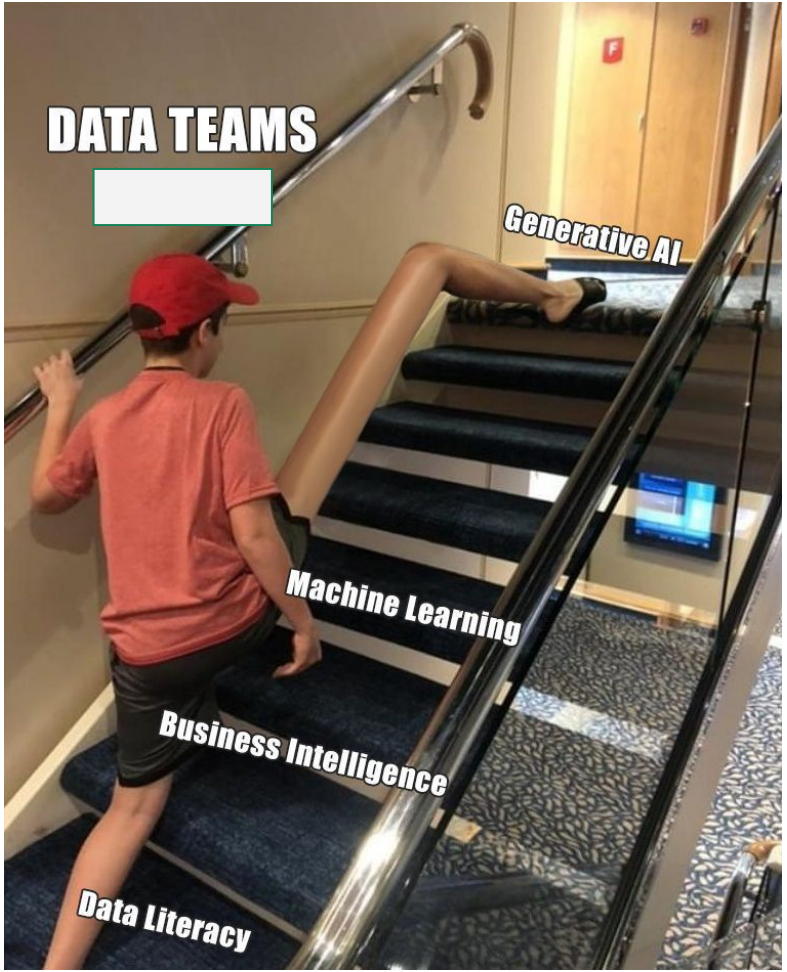
```
if hour < 12:
  print("Good morning!")

def check(number):
  if number > 0:
    return "Positive"
  elif number == 0:
    return "Zero"
  else:
    return "Negative"
```

Traditional Programming


AI
IF
THEN
ELSE
IF
END
THEN
ELSE
END
END
AI


if
tech companies
Is this machine learning?


DATA TEAMS
Generative AI
Machine Learning
Business Intelligence
Data Literacy

bangkit

Going
back to the basics
will strengthen
your
foundations...

Hello to Machine Learning!

bangkit

Say Hello to Machine Learning!

Imagine you are given the task of creating a rule based on this data.

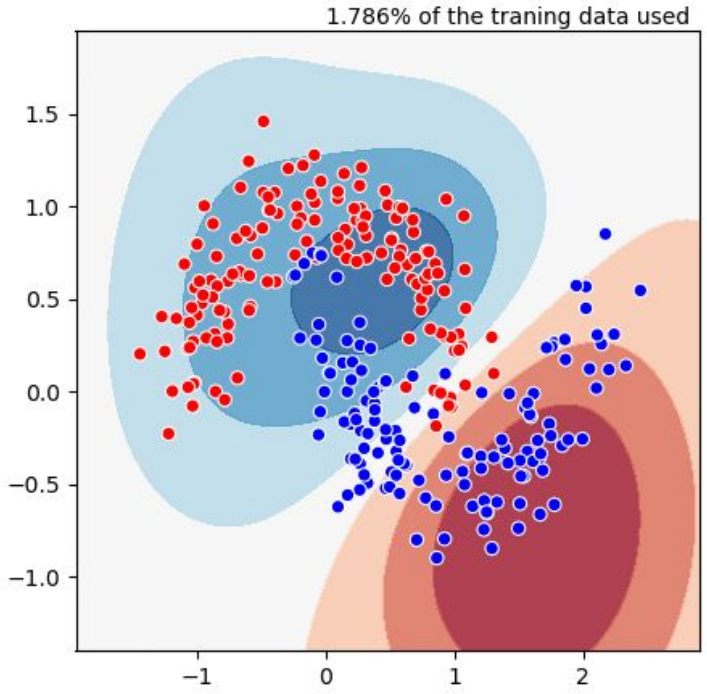
bas	acc	fet	ute	ligh	sev	pro	ab	r	pe	m	hist	hist	hist	hi	hi	hist	hist	hist	hist	hi	f
120.0	0.0	0.0	0.0	0.0	0.0	0.0	73.0	0.5	43.0	2.4	64.0	62.0	126.0	2.0	0.0	120.0	137.0	121.0	73.0	1.0	2.0
132.0	0.006	0.0	0.006	0.003	0.0	0.0	17.0	2.1	0.0	10.4	130.0	68.0	198.0	6.0	1.0	141.0	136.0	140.0	12.0	0.0	1.0
133.0	0.003	0.0	0.008	0.003	0.0	0.0	16.0	2.1	0.0	13.4	130.0	68.0	198.0	5.0	1.0	141.0	135.0	138.0	13.0	0.0	1.0
134.0	0.003	0.0	0.008	0.003	0.0	0.0	16.0	2.4	0.0	23.0	117.0	53.0	170.0	11.0	0.0	137.0	134.0	137.0	13.0	1.0	1.0
132.0	0.007	0.0	0.008	0.0	0.0	0.0	16.0	2.4	0.0	19.9	117.0	53.0	170.0	9.0	0.0	137.0	136.0	138.0	11.0	1.0	1.0
134.0	0.001	0.0	0.01	0.009	0.0	0.002	26.0	5.9	0.0	0.0	150.0	50.0	200.0	5.0	3.0	76.0	107.0	107.0	170.0	0.0	3.0
134.0	0.001	0.0	0.013	0.008	0.0	0.003	29.0	6.3	0.0	0.0	150.0	50.0	200.0	6.0	3.0	71.0	107.0	106.0	215.0	0.0	3.0
122.0	0.0	0.0	0.0	0.0	0.0	0.0	83.0	0.5	6.0	15.6	68.0	62.0	130.0	0.0	0.0	122.0	122.0	123.0	3.0	1.0	3.0
122.0	0.0	0.0	0.002	0.0	0.0	0.0	84.0	0.5	5.0	13.6	68.0	62.0	130.0	0.0	0.0	122.0	122.0	123.0	3.0	1.0	3.0
122.0	0.0	0.0	0.003	0.0	0.0	0.0	86.0	0.3	6.0	10.6	68.0	62.0	130.0	1.0	0.0	122.0	122.0	123.0	1.0	1.0	3.0
151.0	0.0	0.0	0.001	0.001	0.0	0.0	64.0	1.9	9.0	27.6	130.0	56.0	186.0	2.0	0.0	150.0	148.0	151.0	9.0	1.0	2.0
150.0	0.0	0.0	0.001	0.001	0.0	0.0	64.0	2.0	8.0	29.5	130.0	56.0	186.0	5.0	0.0	150.0	148.0	151.0	10.0	1.0	2.0
131.0	0.005	0.072	0.008	0.003	0.0	0.0	28.0	1.4	0.0	12.9	66.0	88.0	154.0	5.0	0.0	135.0	134.0	137.0	7.0	1.0	1.0
131.0	0.009	0.222	0.006	0.002	0.0	0.0	28.0	1.5	0.0	5.4	87.0	71.0	158.0	2.0	0.0	141.0	137.0	141.0	10.0	1.0	1.0
130.0	0.006	0.408	0.004	0.005	0.0	0.001	21.0	2.3	0.0	7.9	107.0	67.0	174.0	7.0	0.0	143.0	125.0	135.0	76.0	0.0	1.0
130.0	0.006	0.38	0.004	0.004	0.0	0.001	19.0	2.3	0.0	8.7	107.0	67.0	174.0	3.0	0.0	134.0	127.0	133.0	43.0	0.0	1.0
130.0	0.006	0.441	0.005	0.005	0.0	0.0	24.0	2.1	0.0	10.9	125.0	53.0	178.0	5.0	0.0	143.0	128.0	138.0	70.0	1.0	1.0
131.0	0.002	0.383	0.003	0.005	0.0	0.002	18.0	2.4	0.0	13.9	107.0	67.0	174.0	5.0	0.0	134.0	125.0	132.0	45.0	0.0	2.0
130.0	0.003	0.451	0.006	0.004	0.0	0.001	23.0	1.9	0.0	8.8	99.0	59.0	158.0	6.0	0.0	133.0	124.0	129.0	36.0	1.0	1.0

bangkit

Say Hello to Machine Learning!

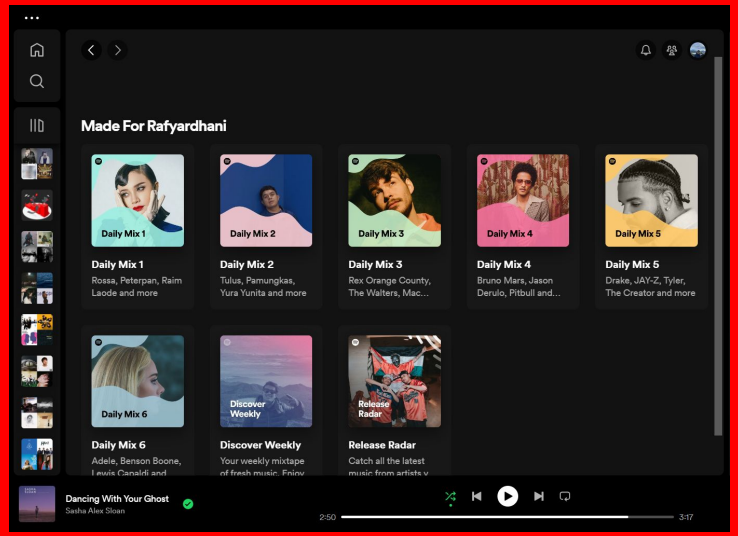
With machine learning, we don't need to create a rules explicit.

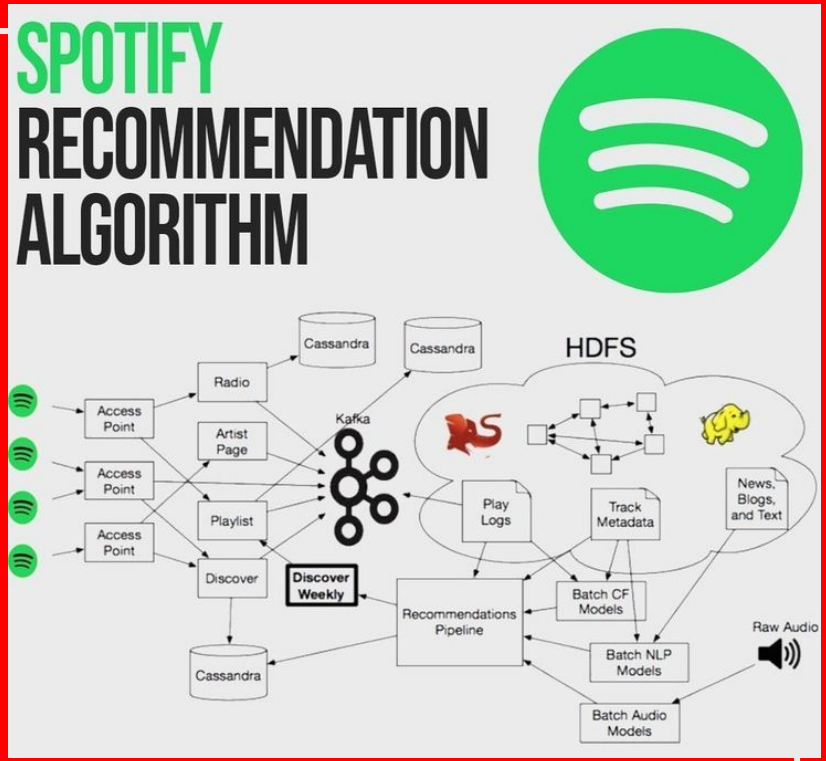
They will find the patterns based on our data.



bangkit

Say Hello to Machine Learning!





bangkit

What is Machine Learning?

“Field of study that **gives computers the ability to learn** without being explicitly programmed.”

Example: Predict Housing Price

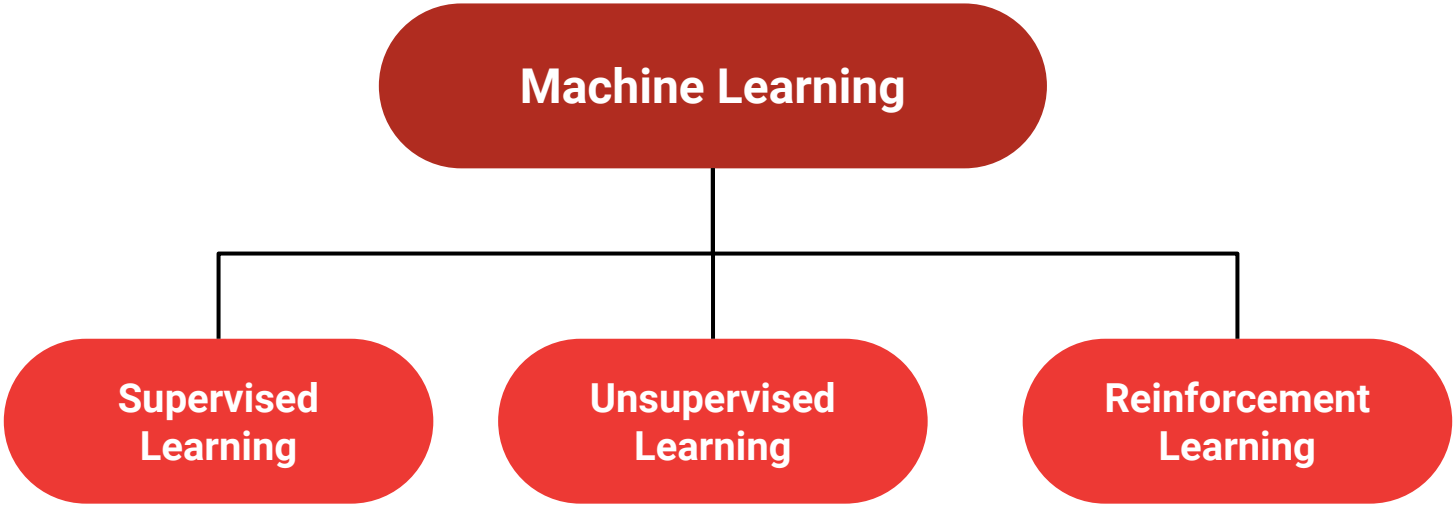
LA	BA	BD	BR	Price
1100	700	5	6	28.000.000.000
824	800	4	4	19.000.000.000
500	400	4	3	4.700.000.000
251	300	5	4	4.900.000.000
1340	575	4	5	28.000.000.000

Why Do We Need Machine Learning?

[redacted]

[redacted]

Types of Machine Learning



Getting Started with Supervised Learning

bangkit

What is Supervised Learning?

- Supervised learning refers to algorithms that **learn how** to map **input X** to **output Y**.
- The critical characteristic of supervised learning is that we give learning algorithm **example data** and **include the right answers** to learn.
- The learning algorithm will **discover the connections** between the example data that produce the correct answers.

bangkit


Data
Answer
Supervised Learning
Rules

Supervised

X_1	X_2	X_p	Y

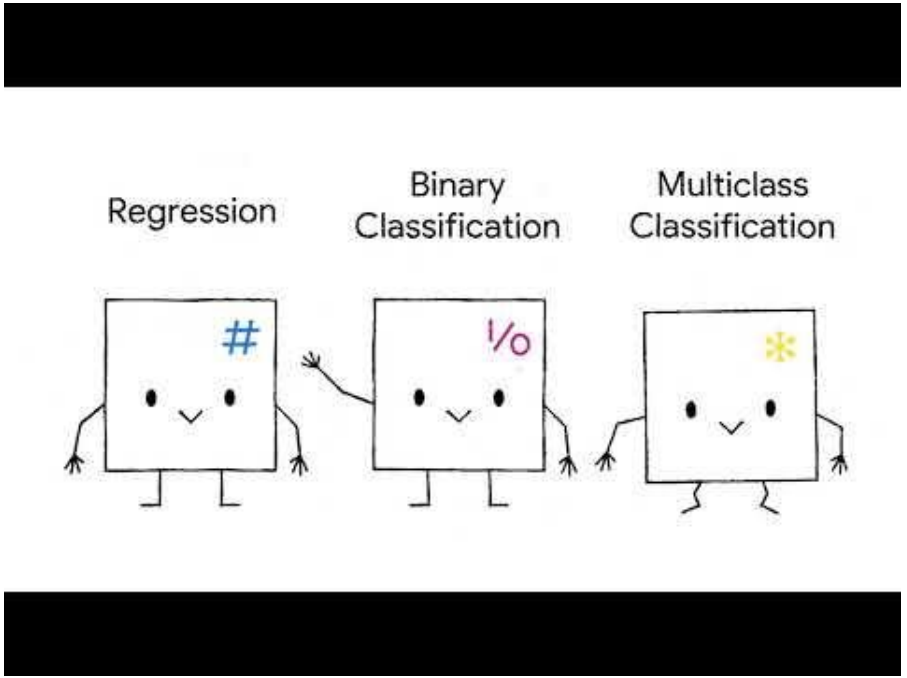
Target

bangkit

Type of Supervised Learning

The most common use cases for supervised learning are **regression** and **classification**:

- Regression model predicts a **numeric value** from **infinitely many possible** numbers.
- Classification model predict the **likelihood** that something belongs to a **category**.



bangkit

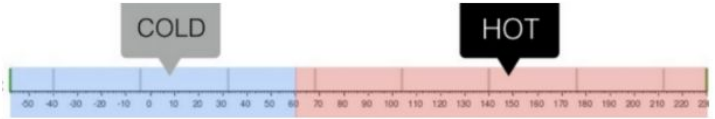
Type of Supervised Learning

Regression

What will be the temperature tomorrow?

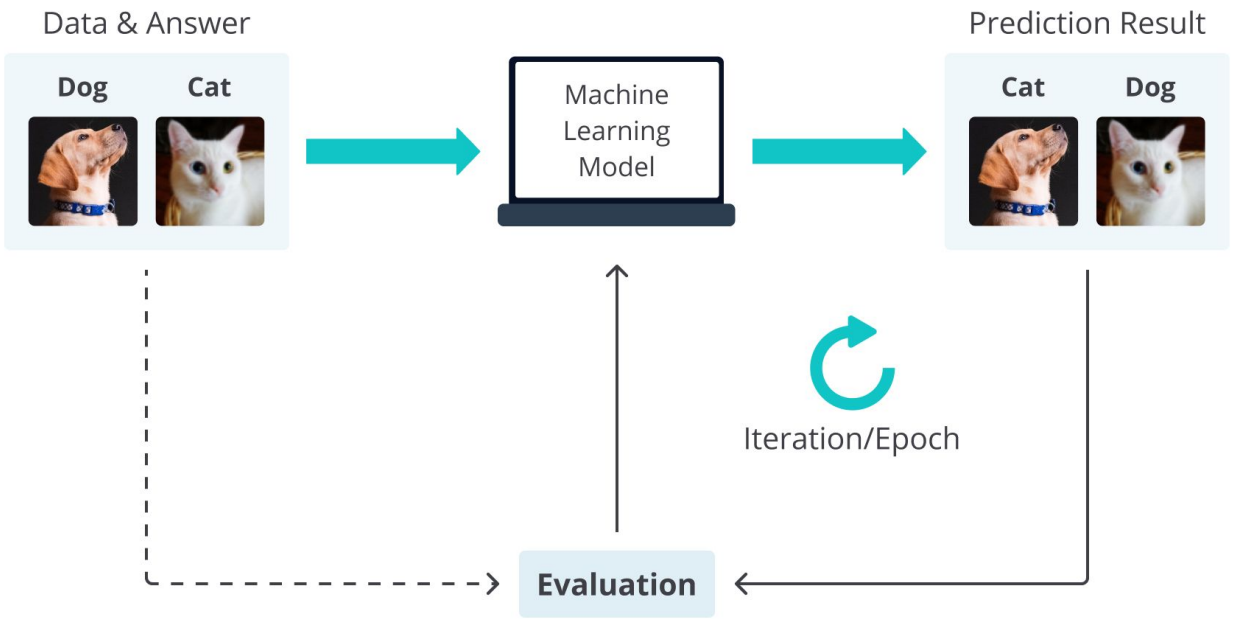
Classification

Will it be hot or cold tomorrow?



bangkit

How Supervised Learning Learn?



Mini Quiz

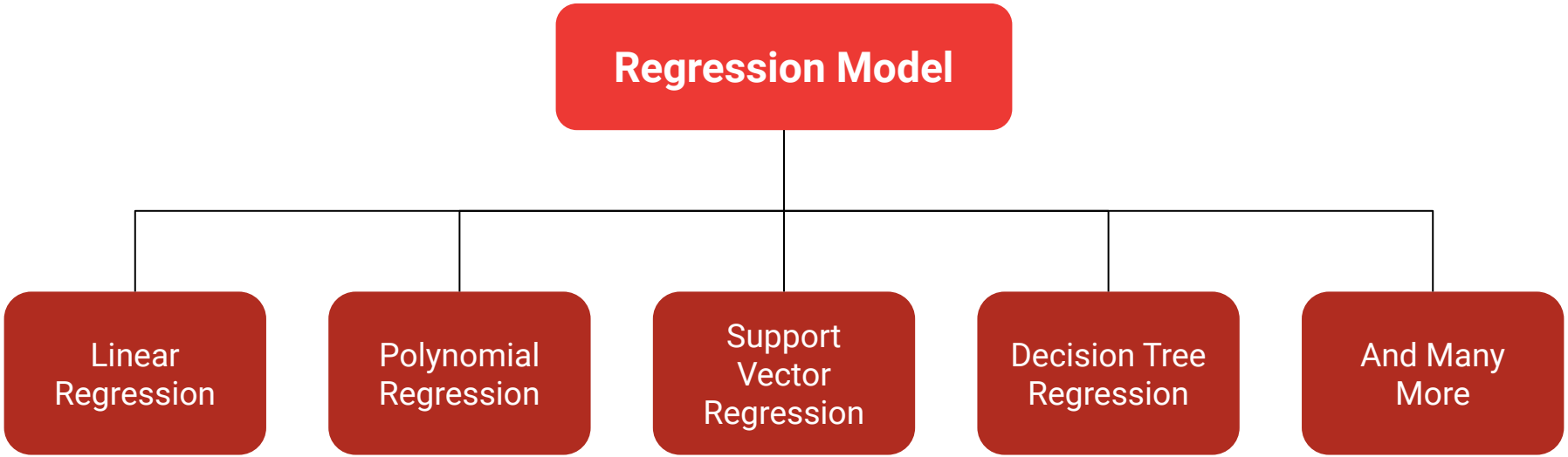
Consider a scenario where you are tasked with predicting the precise amount of rainfall (in millimeters) for the upcoming day in a specific location within a city.

In the context of employing a machine learning algorithm for this prediction, would you categorize this problem as regression or classification, and what factors contribute to your decision?

Regression Model

bangkit

Varieties of Regression Model



Linear Regression Model

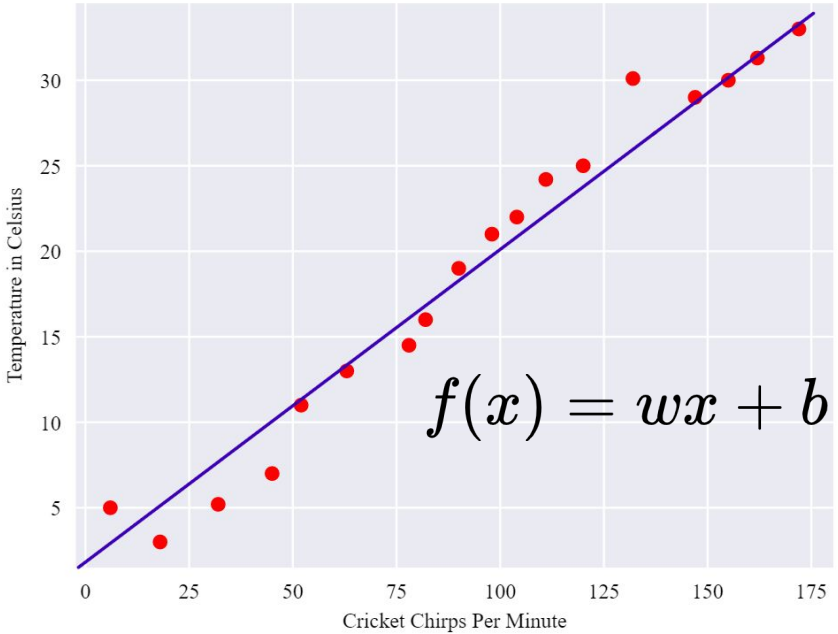
Linear regression is a **linear model** that assumes a **linear relationship** between the **input variables (x)** and the single **output variable (y)**.

Simple linear regression

Multiple linear regression

Understand the Notation

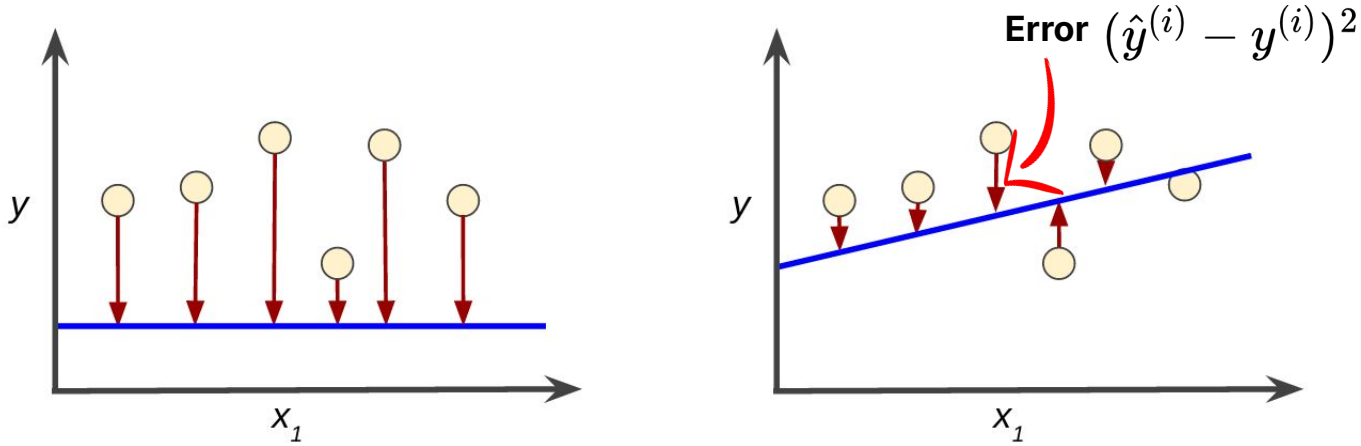
- The x is the **input variabel**.
- The y is the **output or target variabel**.
- The $\hat{y}$ is the **prediction result**.
- The f is the **hypothesis function** to make the predictions.
- The w and b is the **weight** & **bias** parameters.



bangkit

https://developers.google.com/machine-learning/crash-course/descending-into-ml/linear-regression

Cost Function



$$f(x^{(i)}) = \hat{y}^{(i)} = wx^{(i)} + b$$

$$J(w,b) = \frac{1}{2m}\sum_1^m (\hat{y}^{(i)} - y^{(i)})^2$$

bangkit

https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss

Gradient Descent

Gradient descent is an algorithm that you can use to try to **minimize** any function, including the **cost function**.

- Start with some w, b
- **Update the parameters** to reduce the cost function.
- Until we get or near the **global minimum**.

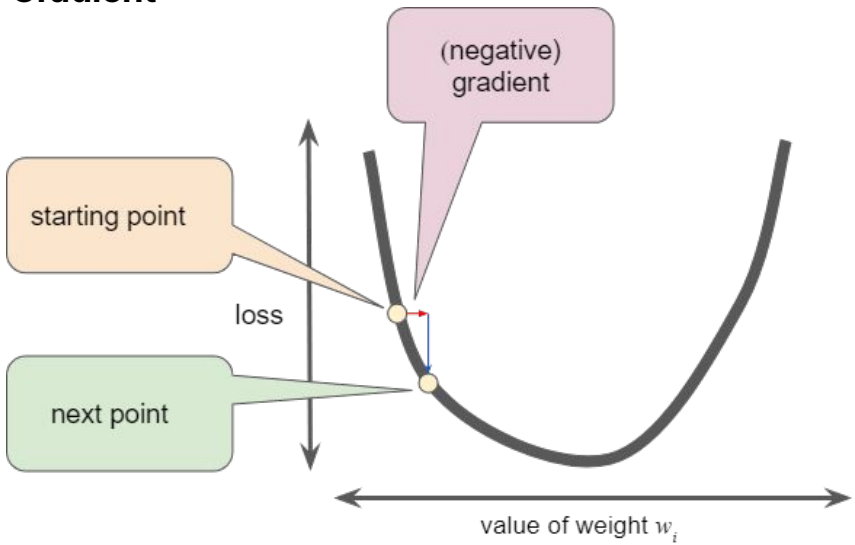
Gradient Descent

$$w_{new} = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

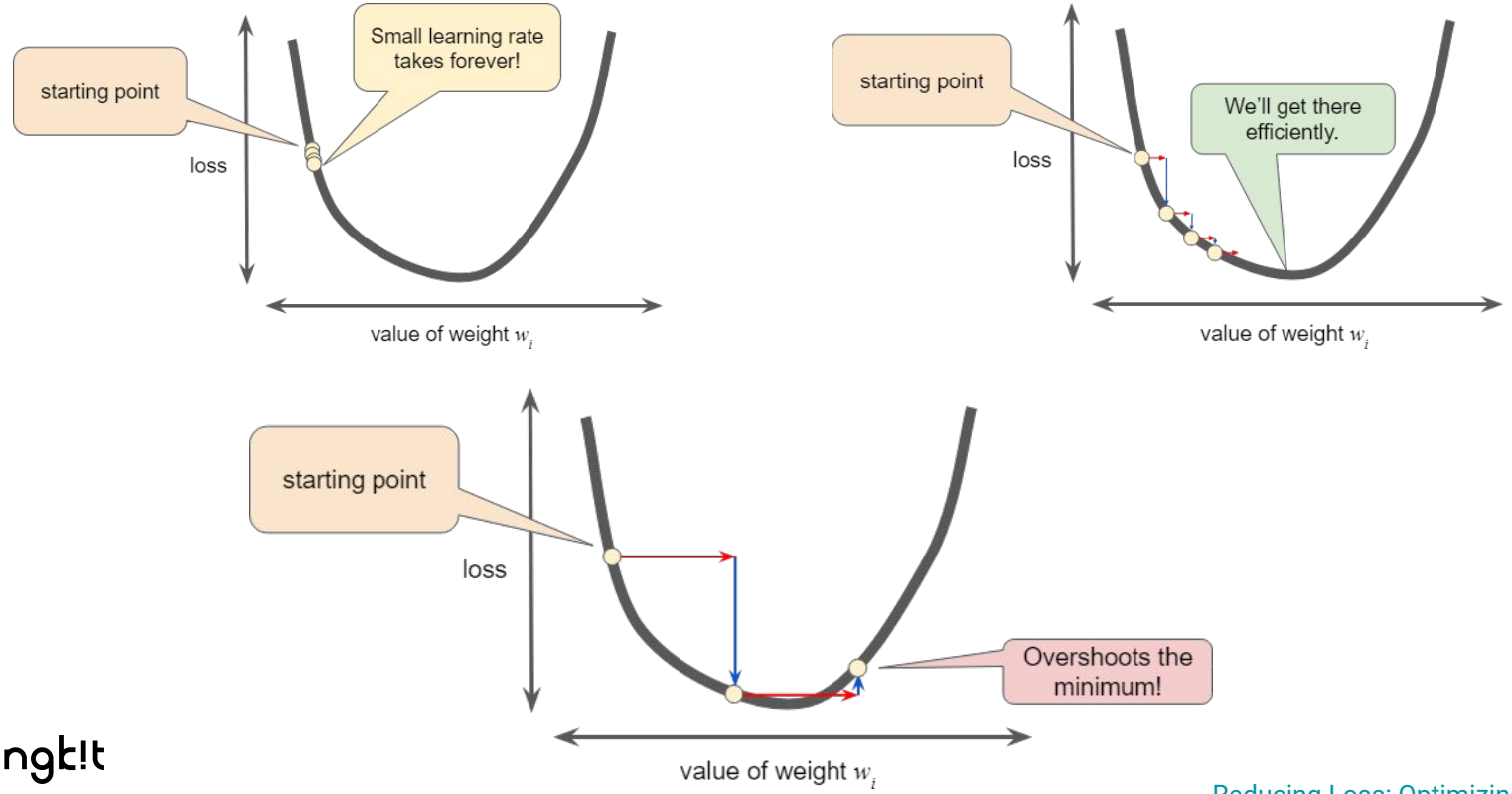
Learning rate

Gradient

$$b_{new} = b - \alpha \frac{\partial}{\partial b} J(w, b)$$



How to Choose the Learning Rate?



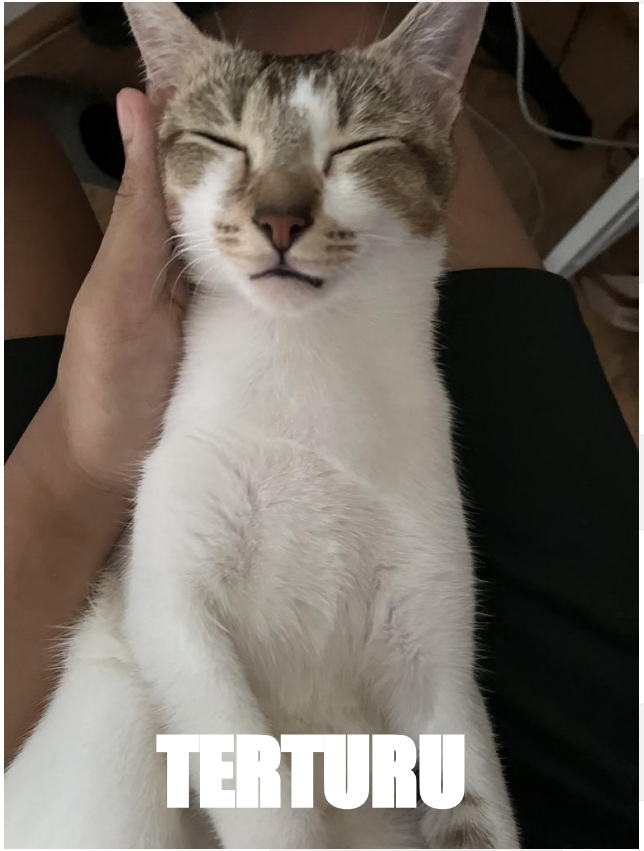
bangkit

Reducing Loss: Optimizing Learning Rate

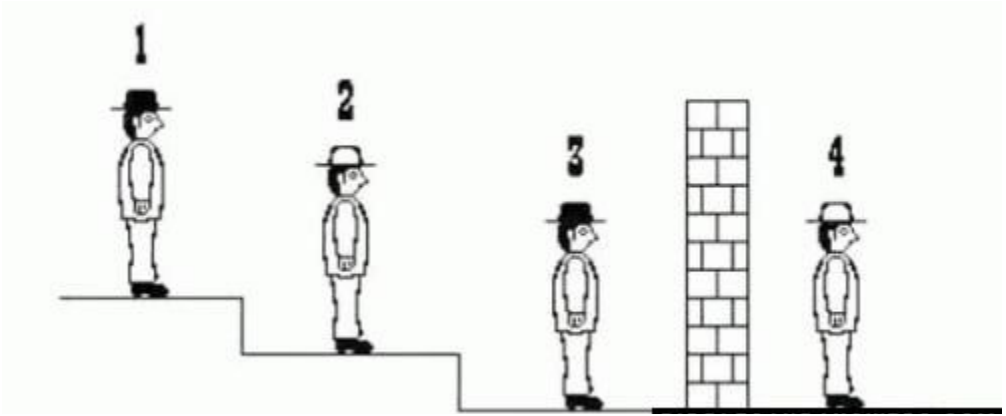
Diving into the World of Regression Algorithms

bangkit

Break Time!



bangkit



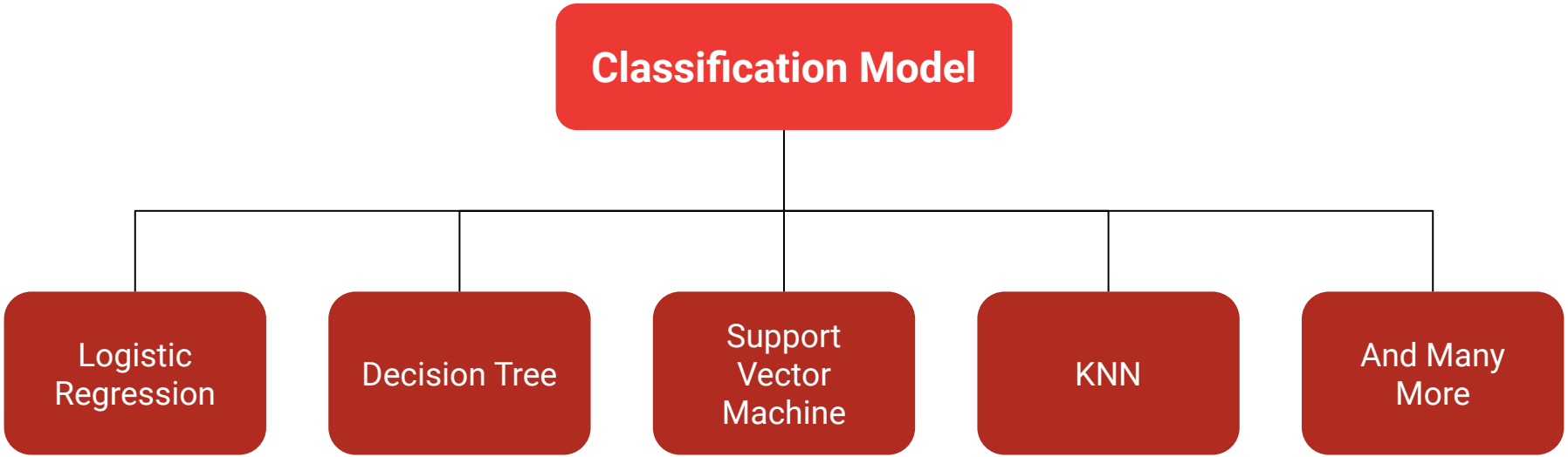
Four Prisoner Hat

Can you solve this riddle?

bangkit

Classification Model

Varieties of Classification Model

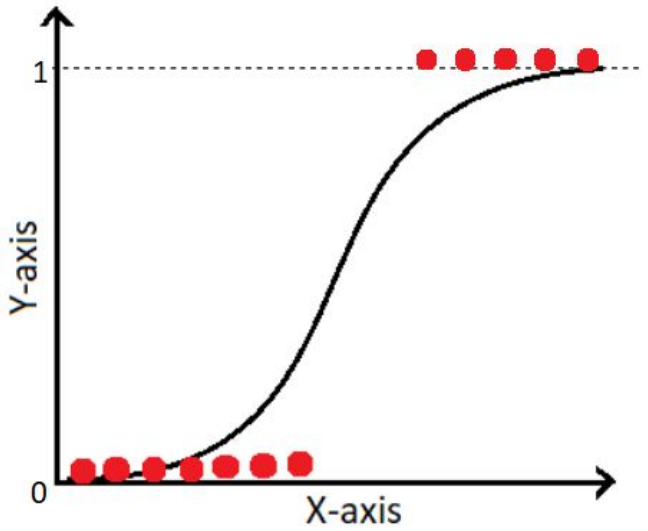


Logistic Regression Model

Logistic regression is a **classification algorithm** that is used to **predict the probability of a categorical** dependent variable.

Typically used to model a **binary classification**

Used **sigmoid function**

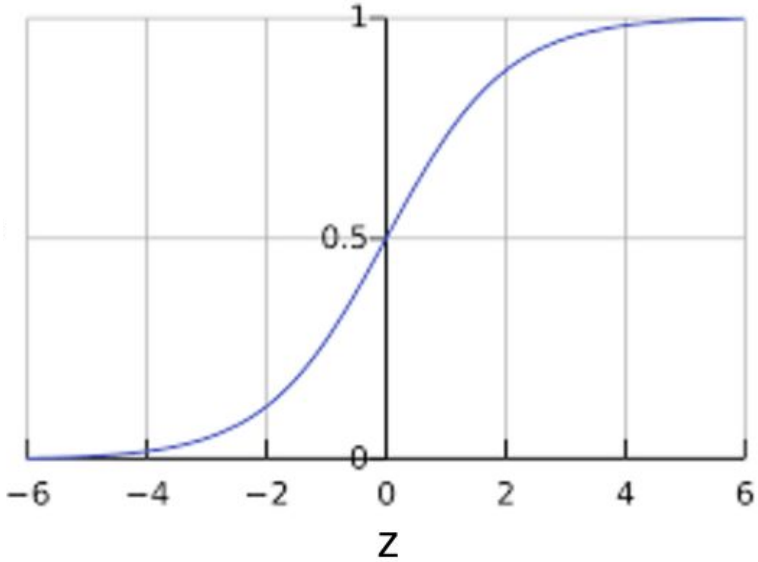


Sigmoid Function

- Sigmoid is a mathematical function that takes **any real number** and **maps it to** a probability **between *1* and *0***.
- This function assists the logistic regression model to **squeeze the values from (*-n*,*+n*) to (*0*,*1*)**

Sigmoid Function

$$f(x) = wx + b = z \quad g(z) = \frac{1}{1+e^{-z}}$$



$$0 < g(z) < 1$$

Evaluation Metrics for Classification

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}}$$

The proportion of correct predictions to the total predictions made.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The proportion of positive data that is correctly predicted by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The proportion of correct positive predictions compared to total positive predictions.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Harmonic mean of precision and recall, useful when there is an imbalance between classes.

Classify with Style:
An Exciting Voyage through Algorithmic Waters

bangkit

Other Important Things

Feature Scaling

Feature scaling is a method used to **normalize the range of independent** variables or features of data.

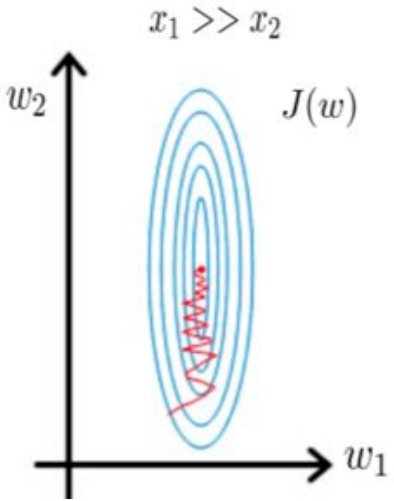
- **Min-Max scaling** is a scaling technique in which values are **shifted and rescaled between 0 and 1**.
- **Z-score normalization** is scaling technique where the values are **centered around the mean with a unit standard deviation**.

Feature Scaling

$$x' = \frac{x - x_{max}}{x_{max} - x_{min}}$$

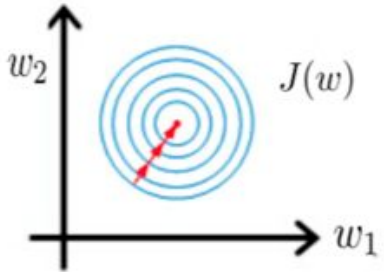
$$x' = \frac{x - \mu}{\sigma}$$

Gradient descent without scaling



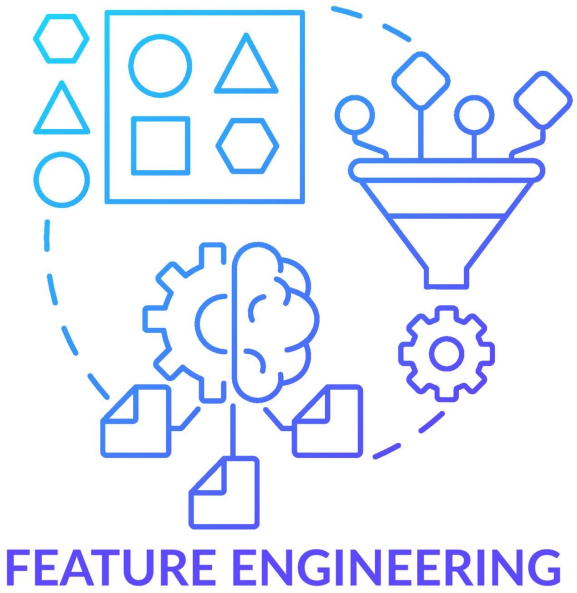
Gradient descent after scaling variables

$0 \le x_1 \le 1$
$0 \le x_2 \le 1$

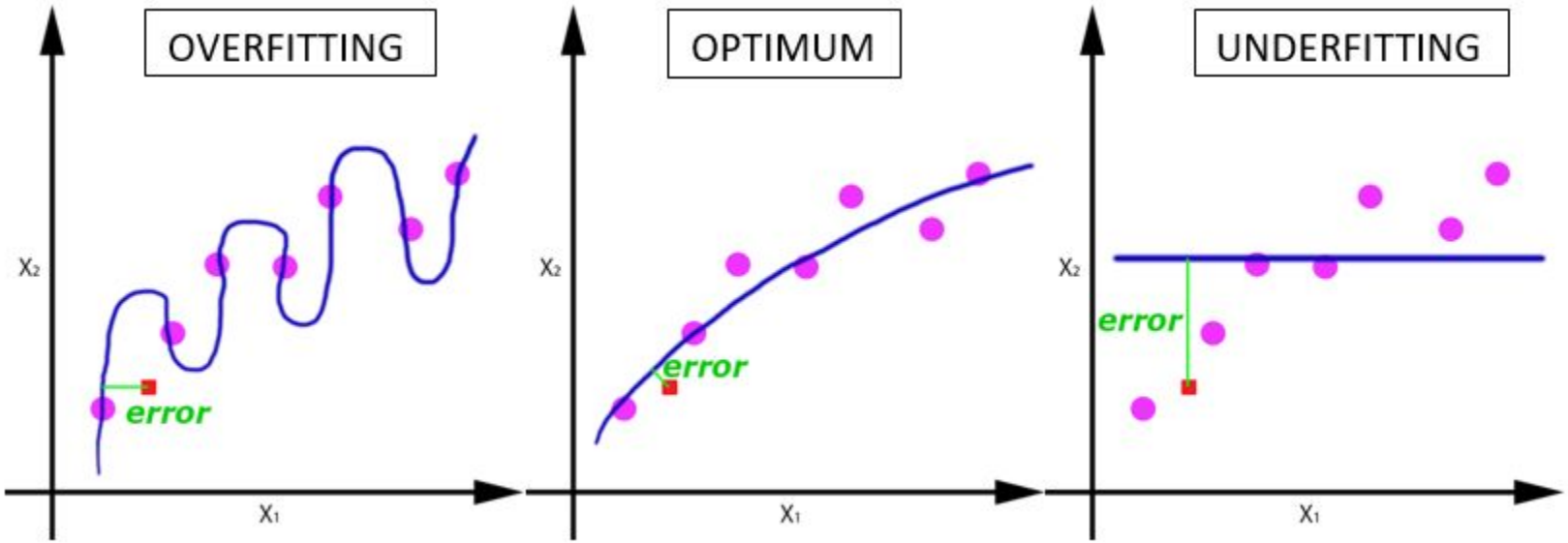


Feature Engineering

- Feature engineering is a technique that **leverages data** to **create a new feature** that isn't in the training set.
- The goal is to **simplify** and **speed** up the training process while **enhancing model accuracy**.



Overfitting & Underfitting

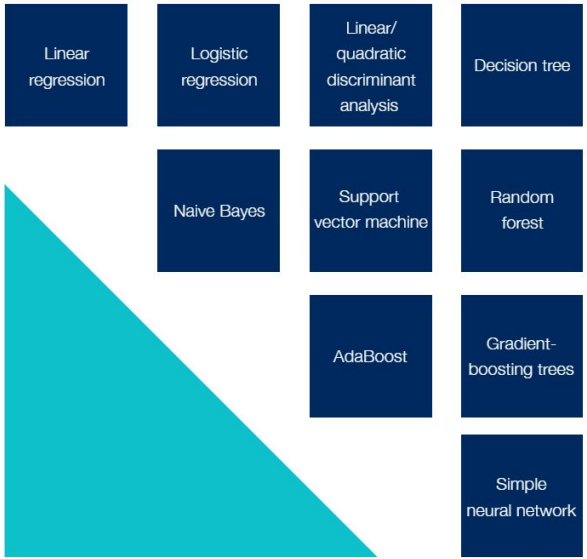


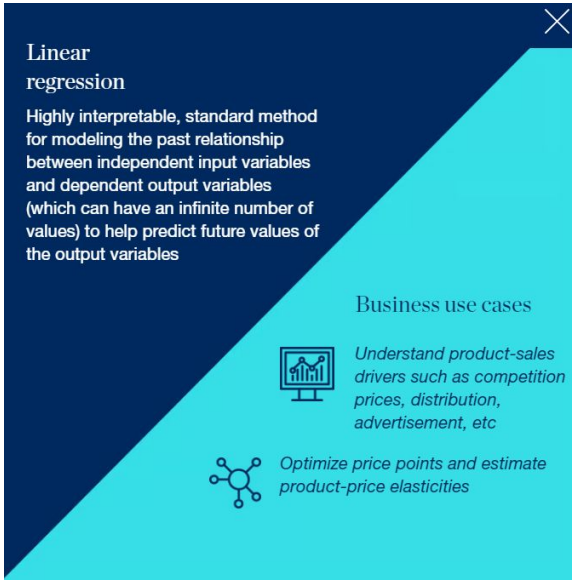
bangkit

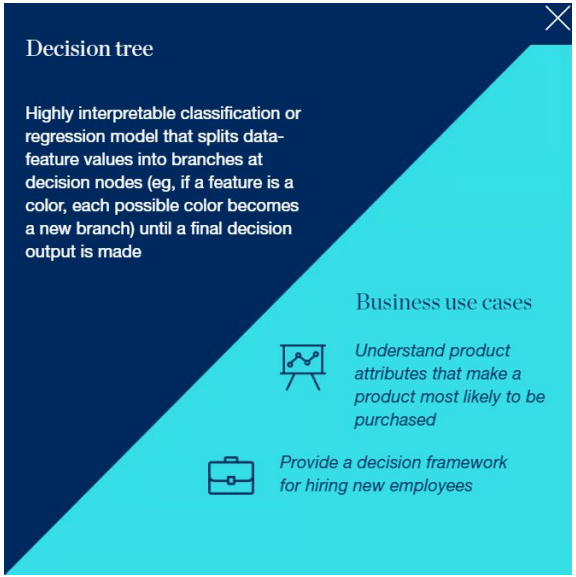
Industry Impact/Example

https://www.mckinsey.com/capabilities/quantumblack/our-insights/an-executives-guide-to-ai

Supervised Learning







bangkit

Industry Impact/Example

https://www.mckinsey.com/capabilities/quantumblack/our-insights/an-executives-guide-to-ai

Reinforcement Learning

Optimize the trading strategy for an options-trading portfolio

Balance the load of electricity grids in varying demand cycles

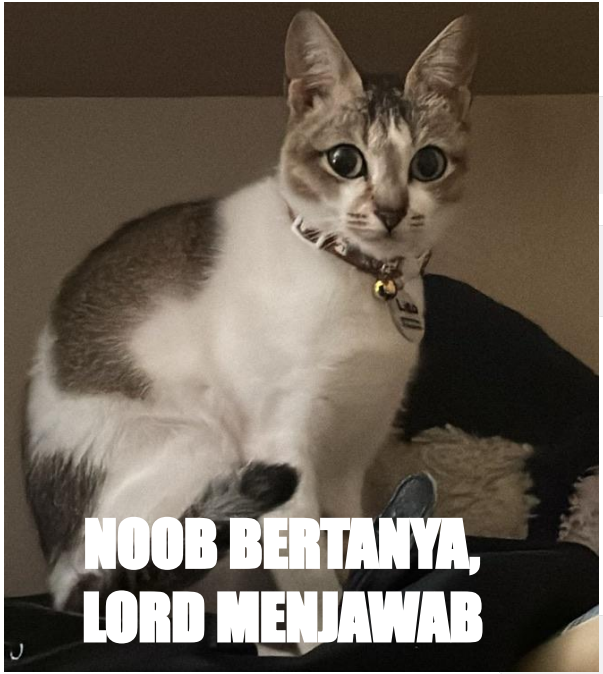
Stock and pick inventory using robots

Optimize the driving behavior of self-driving cars

Optimize pricing in real time for an online auction of a product with limited supply

bangkit

Discussions



Quiz and Feedback

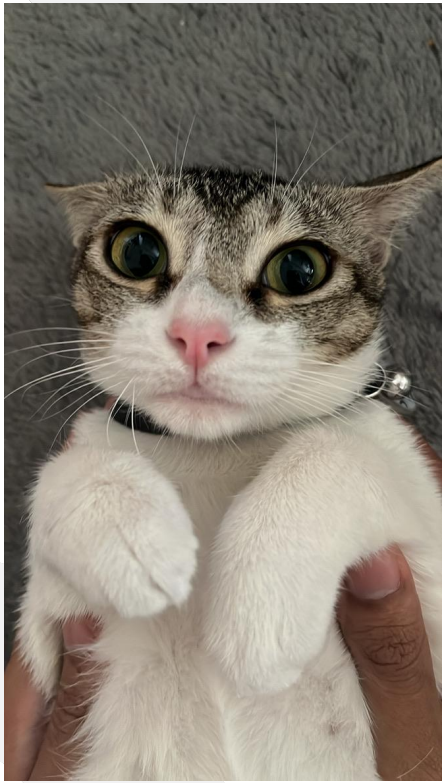
bangkit

Conclusion

You have learned **basic concepts in machine learning, how regression and classification models work**

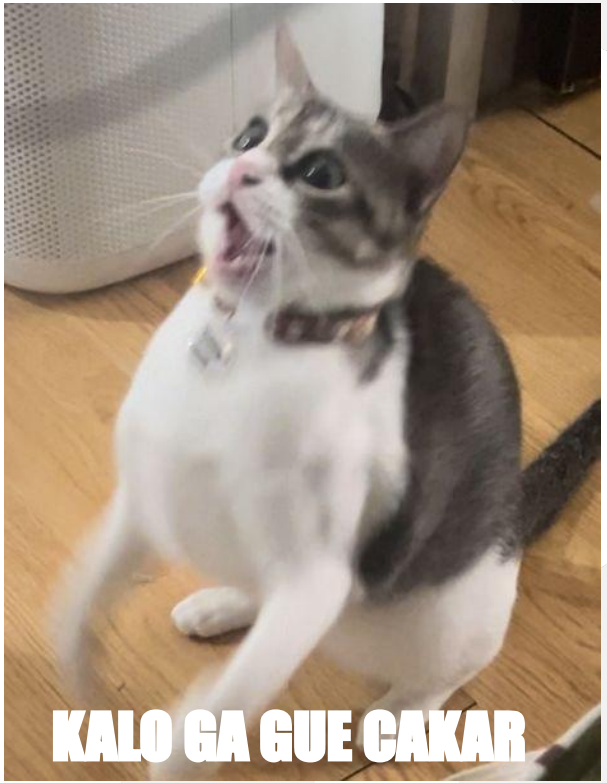
These knowledges will give you a solid fundamental concept to **becoming a machine learning practitioner**

Thank You



Follow majikan saia:
Tiktok: @jenniethecatwoman

Currently at ~1500 Followers



bangkit