AI, Machine Learning, and Deep Learning

Sudong Lee

sudonglee@ulsan.ac.kr

https://dais.ulsan.ac.kr/

Goal

"Define AI, machine learning, and deep learning and understand their relationship."

Contents

- Definition of AI
- Definition of machine learning
- Machine learning pipeline
- Definition of deep learning

Definition of AI

Definition of AI in One Sentence

"Can you define AI in one sentence?"

Richard Phillips Feynman

The esteemed scientist *Richard Feynman* once said,

"What I cannot create, I do not understand."

In response, I say,

"What I cannot define in one sentence, I do not understand."

Definition of AI as an Image

"What image comes to your mind with AI?"

Literally Speaking…

"Artificial Intelligence = Artificial + Intelligence"

Created by human beings

Artificial Intelligence

The ability to think, reason, and understand
instead of doing things automatically or by instinct.

What is *Intelligence*?

"The human brain processes an input to produce a required output."

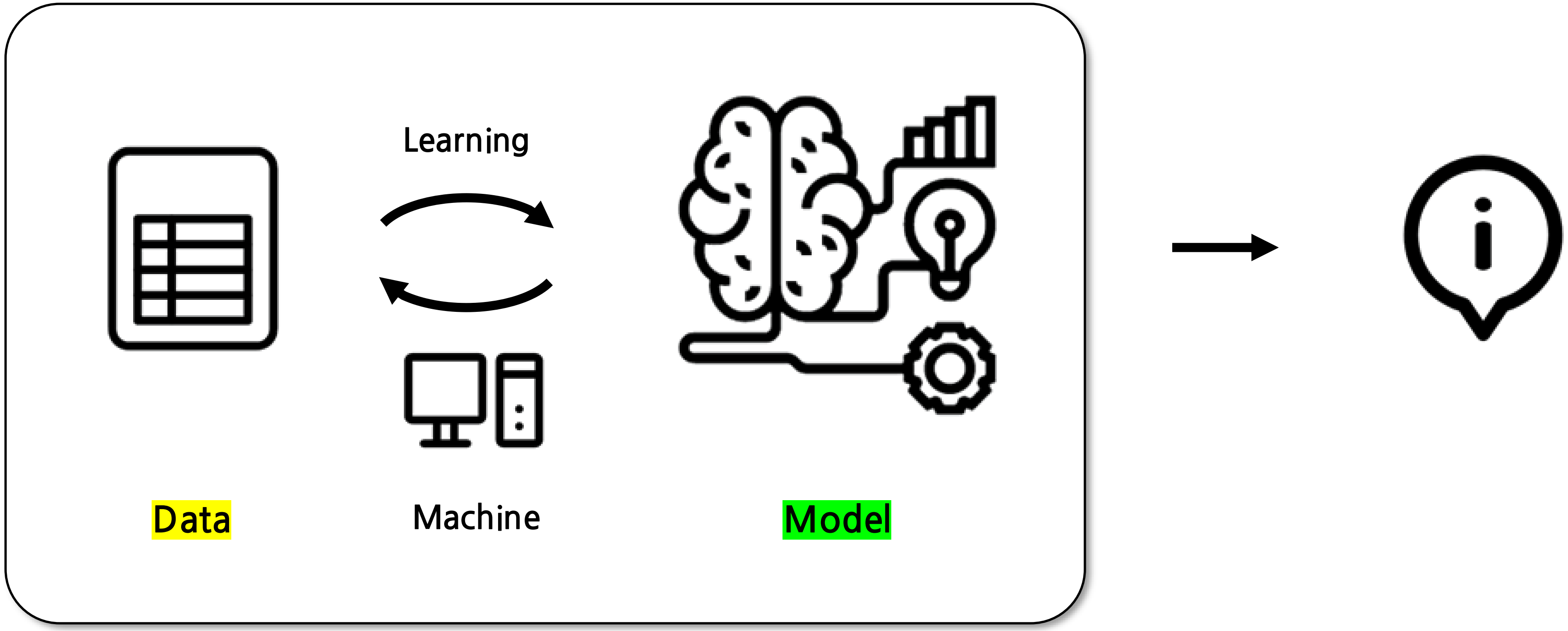
"Dog 🐶" or "Cat 🐱"

"Adorable! 😍" or "Run!! 😨"

Definition of Machine Learning

Definition of Machine Learning

"AI techniques that enable computers to automatically learn from data to produce a required output for their user."



Machine Learning

A Simple Example of Machine Learning

X	Y
1	2
2	4
3	6
4	8

Data

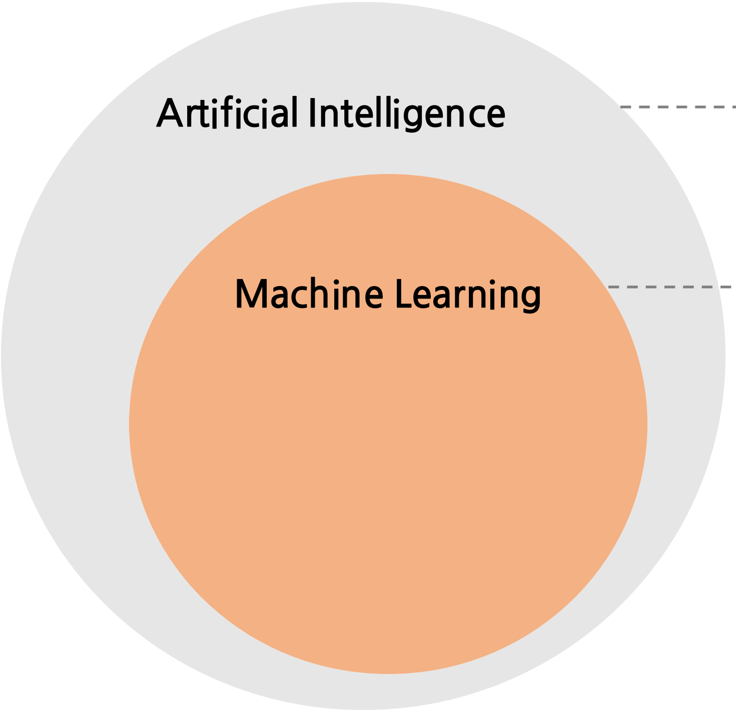
Problem $X = 5 \longrightarrow Y = ?$

Model $Y = wX$

Learning

$w = 2$

AI, Machine Learning, and Deep Learning



Any technique that enables computers to mimic intelligence

AI techniques that enable computers to automatically learn from data to generate the desired output for users

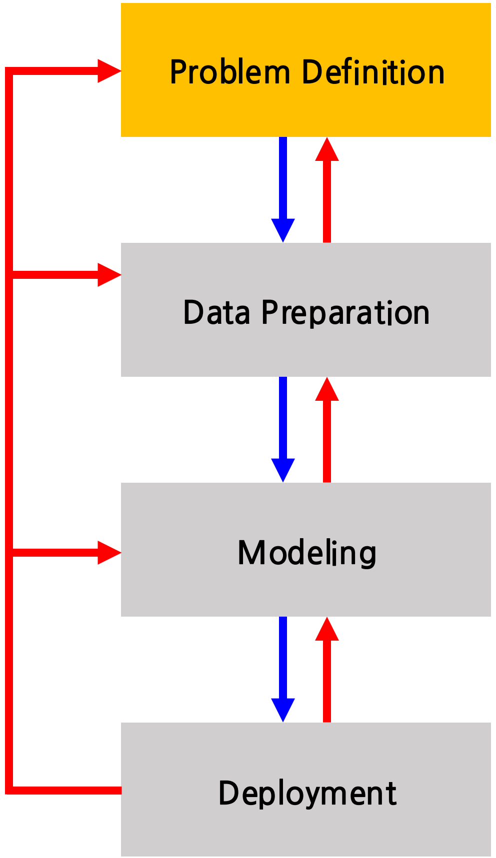
"Is this AI? ML?"

"The air conditioner automatically turns on/off if the room temperature is above/below 25℃."

Scenario	AI	ML
The user has set the temperature to 25℃.	O	X
The air conditioner learned the optimal temperature, 25℃, from past user behaviors.	O	O

Machine Learning Pipeline

Machine Learning Pipeline



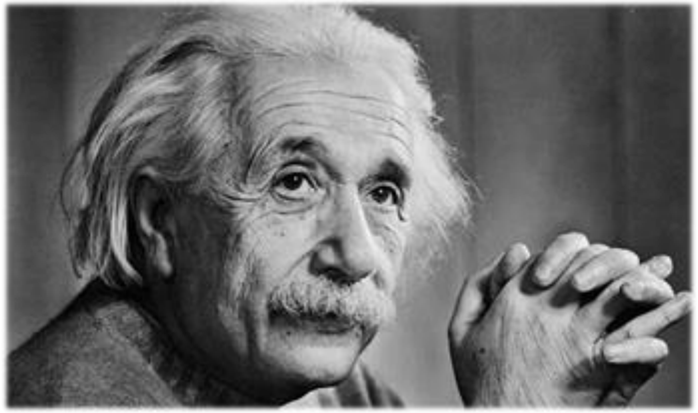
- Define a goal and problem type.
- Define data (X and Y).

- Data collection
- Data preprocessing
- Data splitting

- Learning
- Evaluation

- Test (Inference on new data)
- Monitor and feedback

Importance of Problem Definition



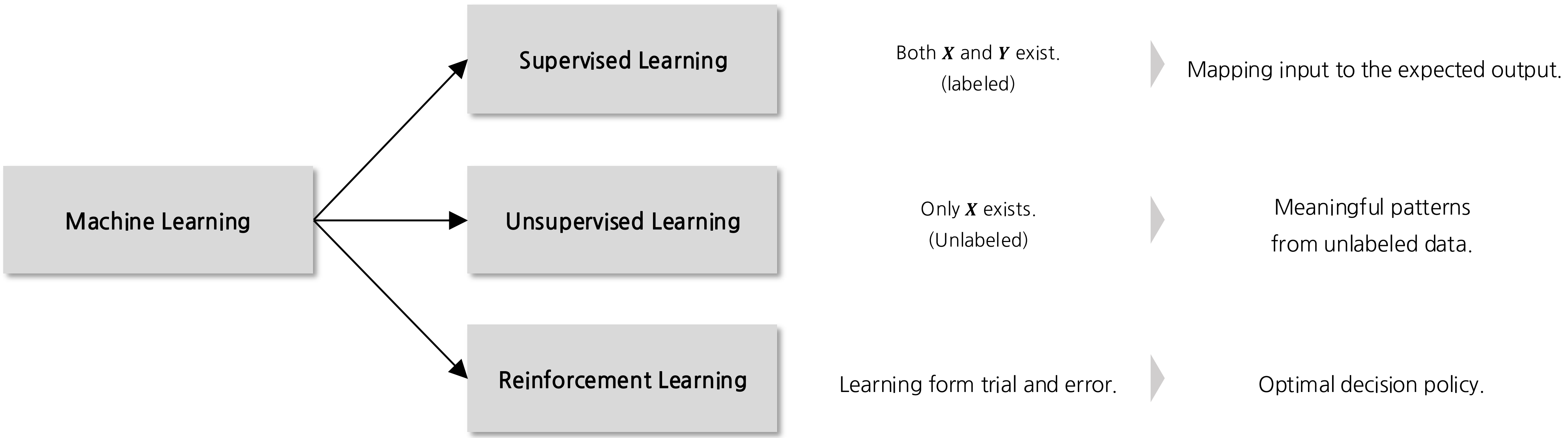
"If I had an hour to solve a problem

I'd spend 55 minutes thinking about the problem

and five minutes thinking about solutions."

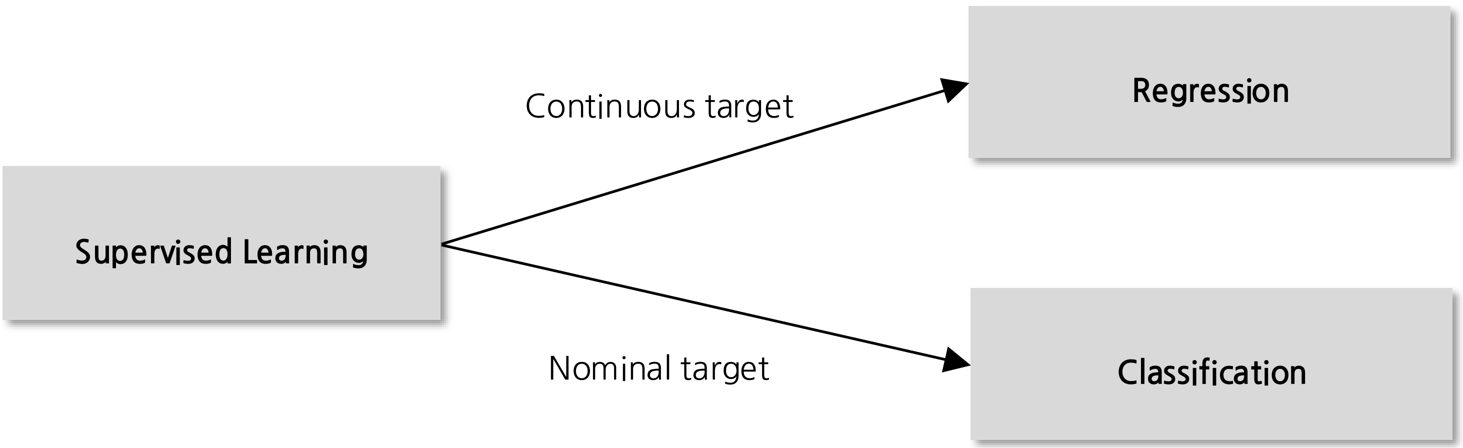
Albert Einstein

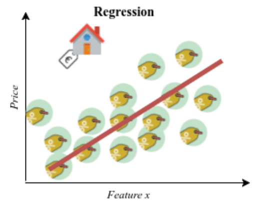
Primary Learning Paradigms in Machine Learning

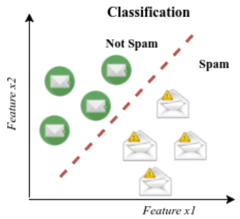


Supervised Learning

“Learning a function that maps an input to an output based on example input-output pairs.”

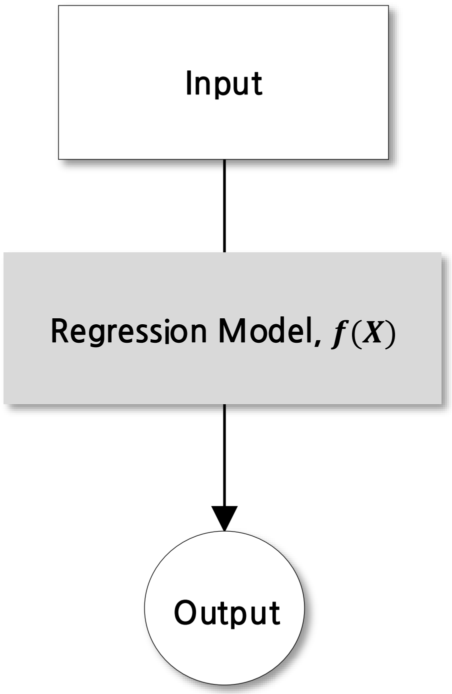


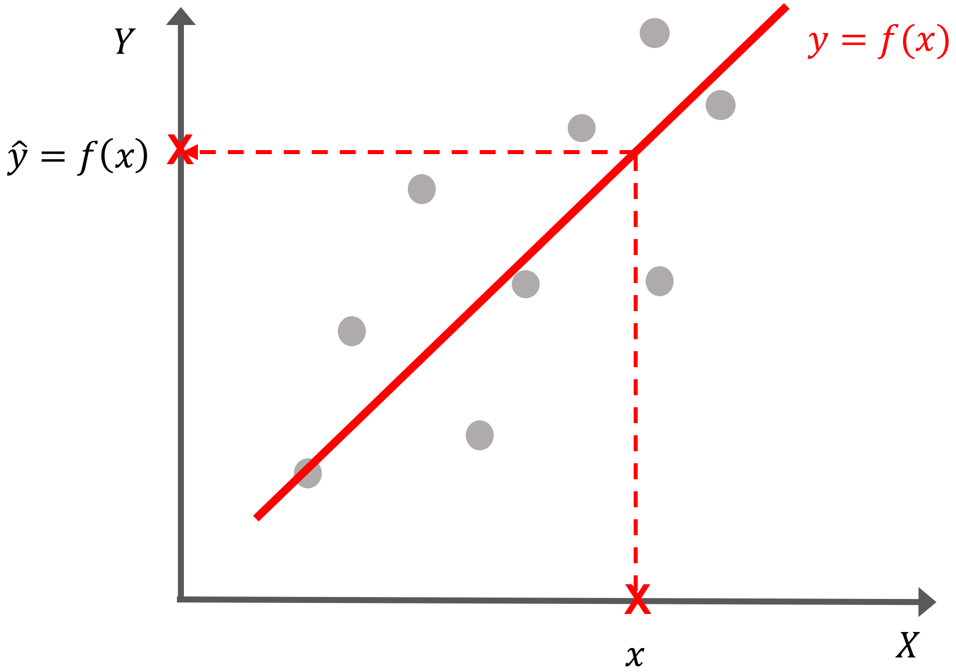




Regression

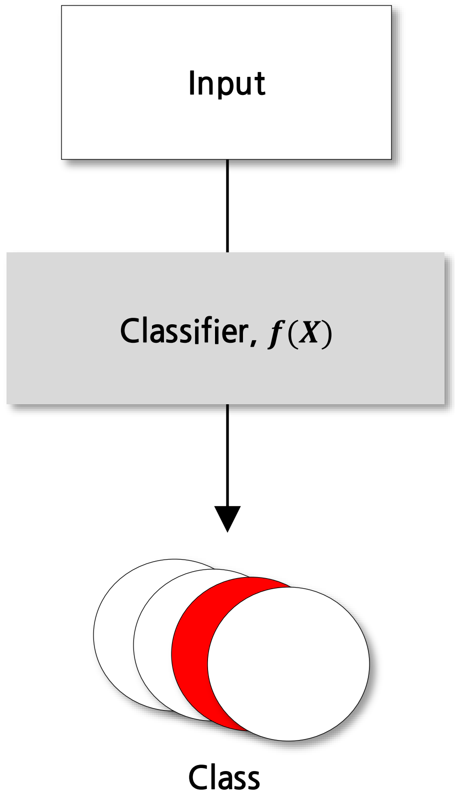
"Predict a continuous output value based on one or more input feature values."

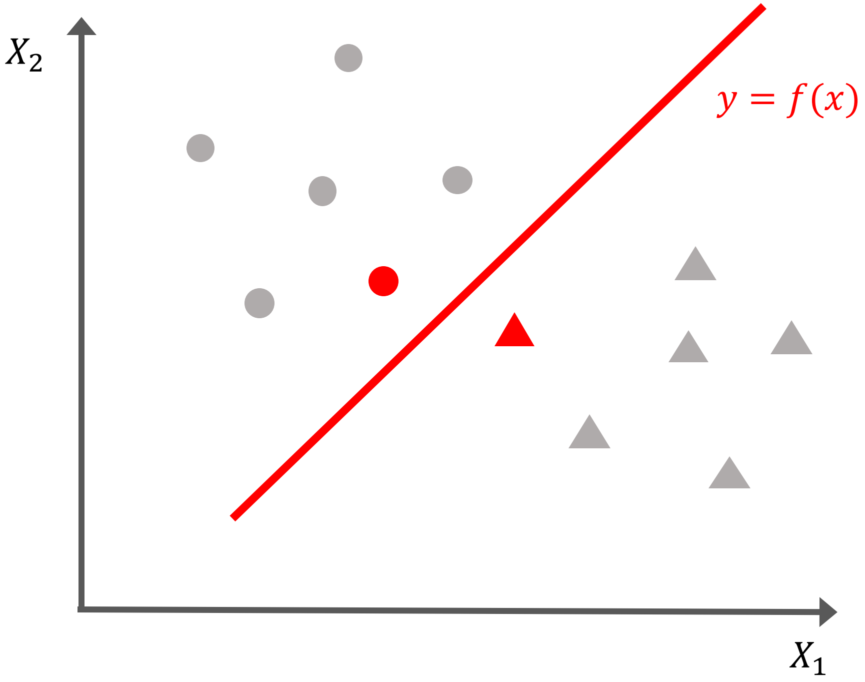




Classification

"Classify input data into two or more categorical classes."



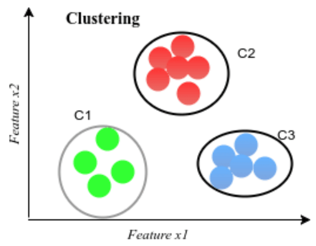


Unsupervised Learning

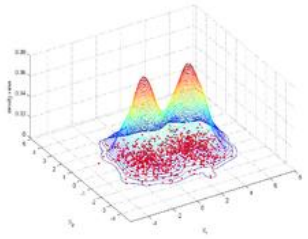
"Find meaningful patterns in the data itself, not in the relationship between input and output."

Unsupervised Learning

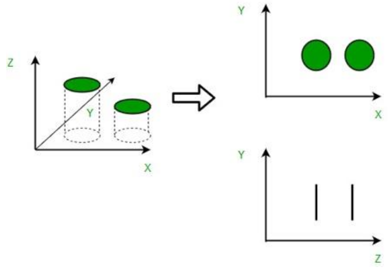
Discover groups of similar observations → **Clustering**



Determine the distribution of data → **Density Estimation**

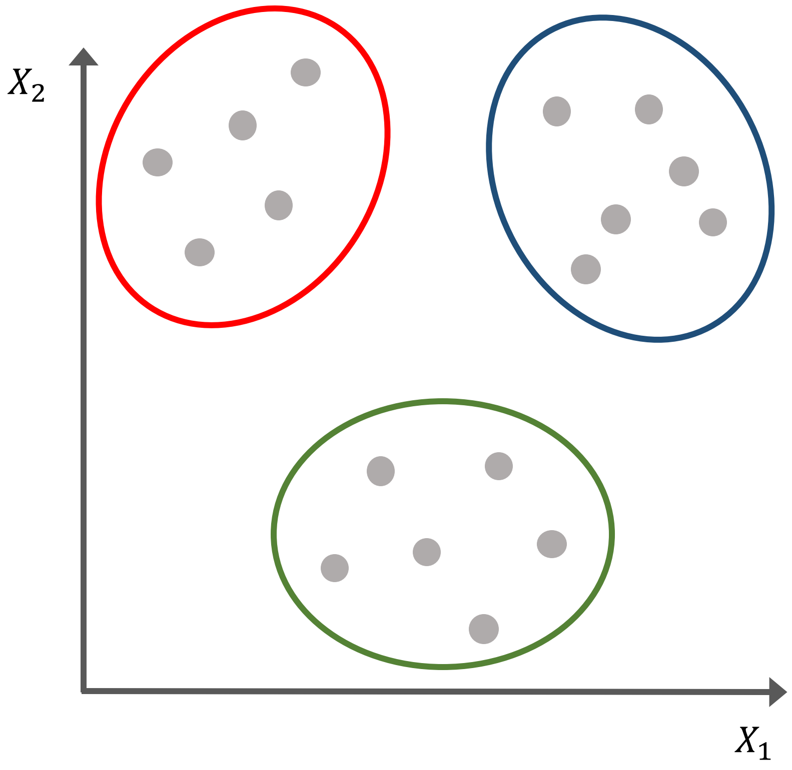


Project the data from a high-dimensional space down to lower dimensions → **Dimensionality Reduction**



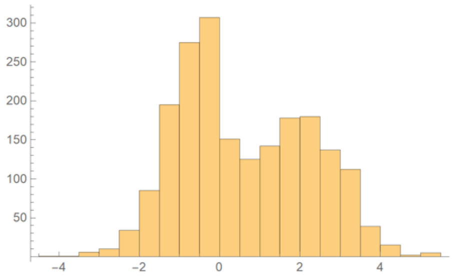
Clustering

"Identifying homogeneous subgroups among the observations"

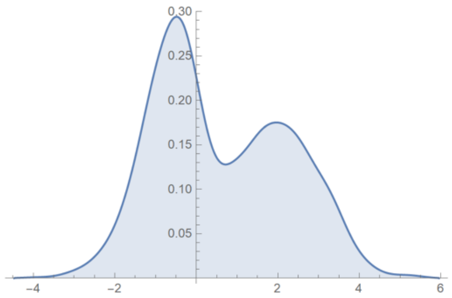


Density Estimation

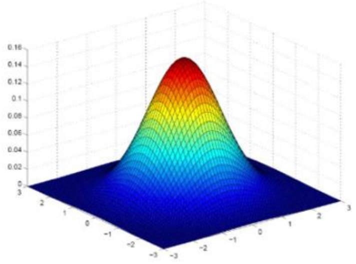
"Estimate the probability density of data based on observations."



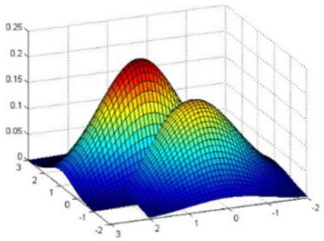
Histogram



Probability Density Function (PDF)

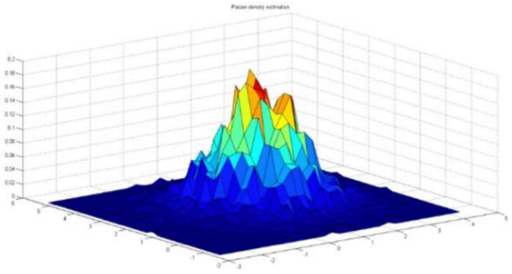


Gaussian Density Estimation



Mixture of Gaussian

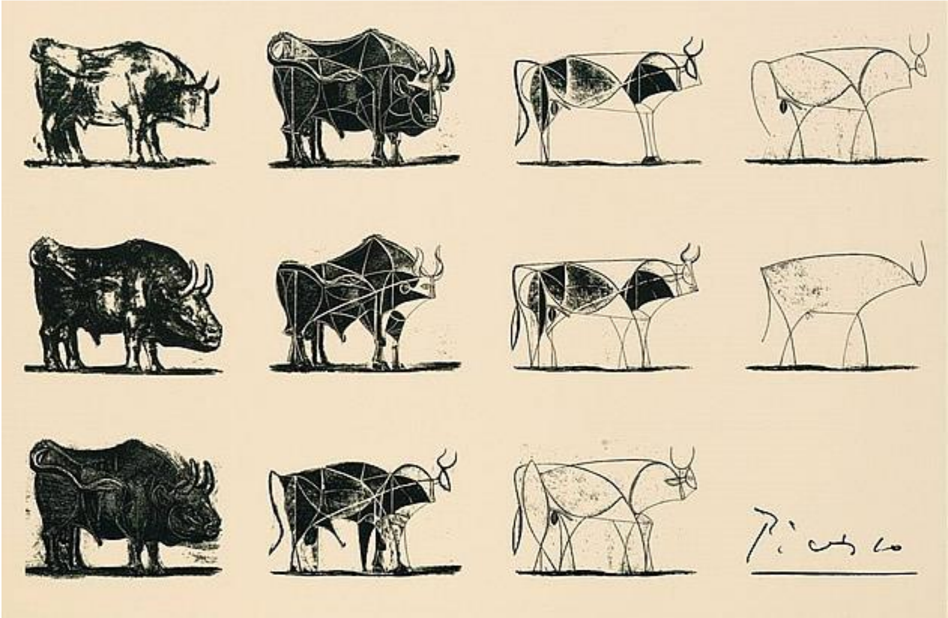
Density Estimation



Kernel Density Estimation

Dimensionality Reduction

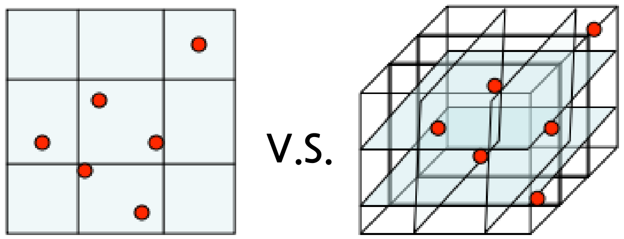
"Transformation of data from a high-dimensional space into a low-dimensional space while minimizing the loss of information and retaining meaningful properties of the original data."

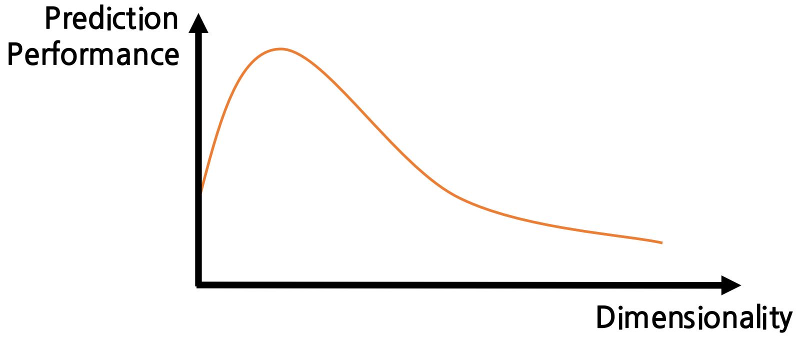


Picasso's bulls

Dimensionality Reduction

- The curse of dimensionality
  - For a data given size,

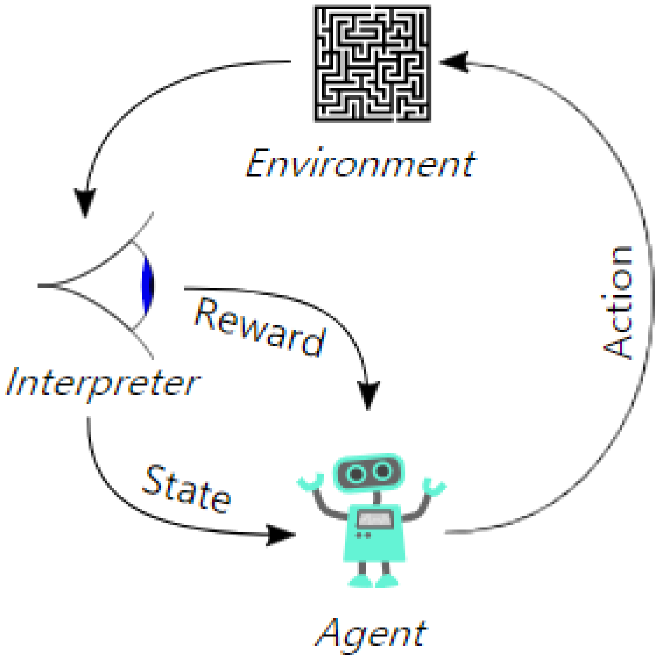




- Advantages of dimensionality reduction
  - Decrease in prediction error
  - Shorter training time
  - Smaller number of training data required
  - Improved model interpretability
  - Enhanced generalization by reducing overfitting

Reinforcement Learning

"Find suitable actions to take in a given situation in order to maximize a reward."

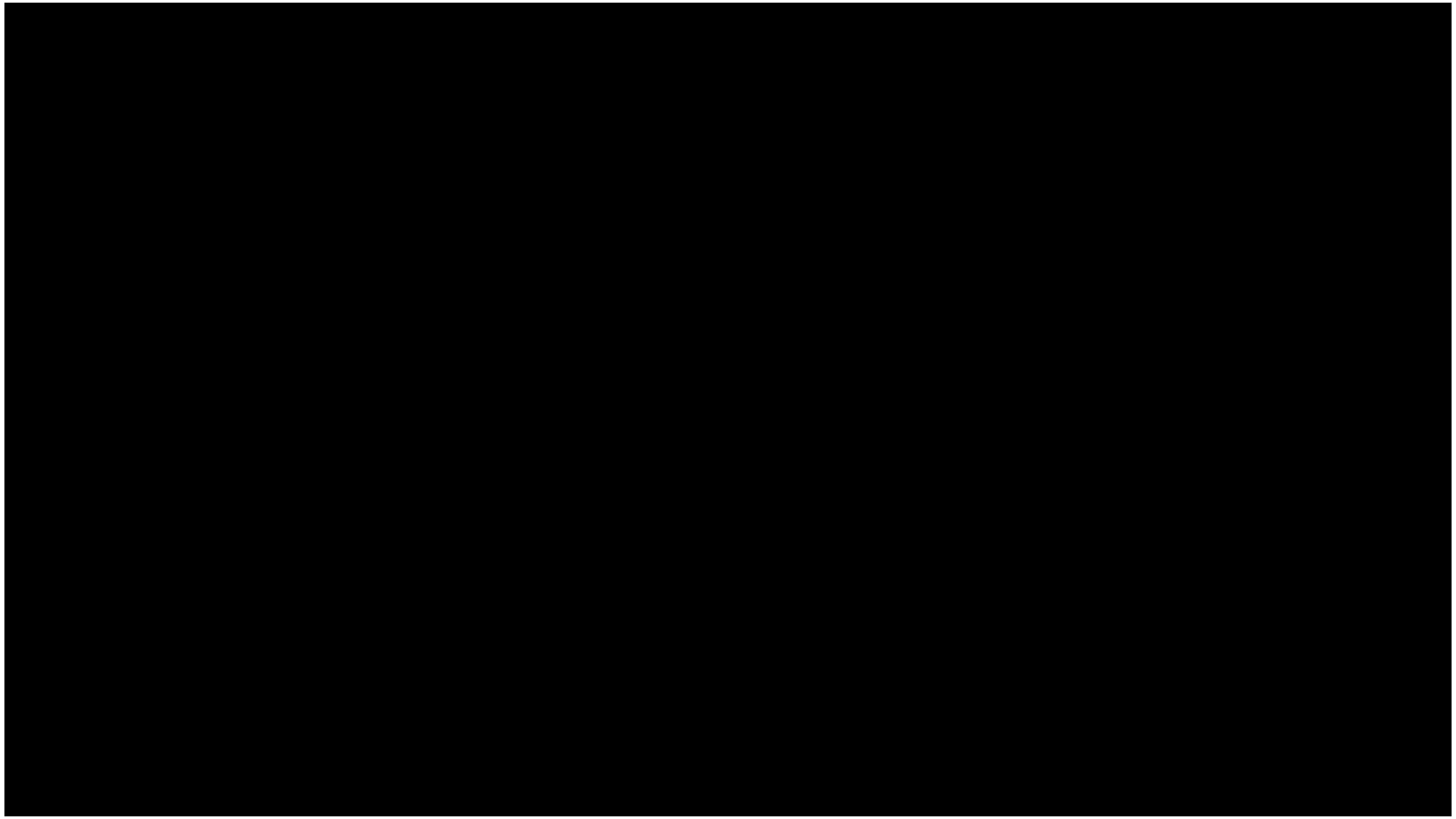


(Image Source | Wikipedia)

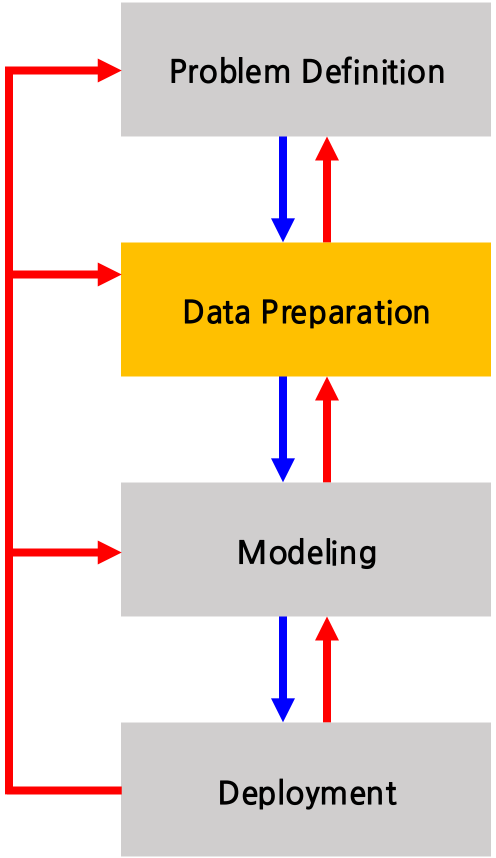
Sedol Lee and AlphaGo's Go match in 2016

(Image Source | 한국일보)

“Teaching a robot to walk using reinforcement learning”



Machine Learning Pipeline



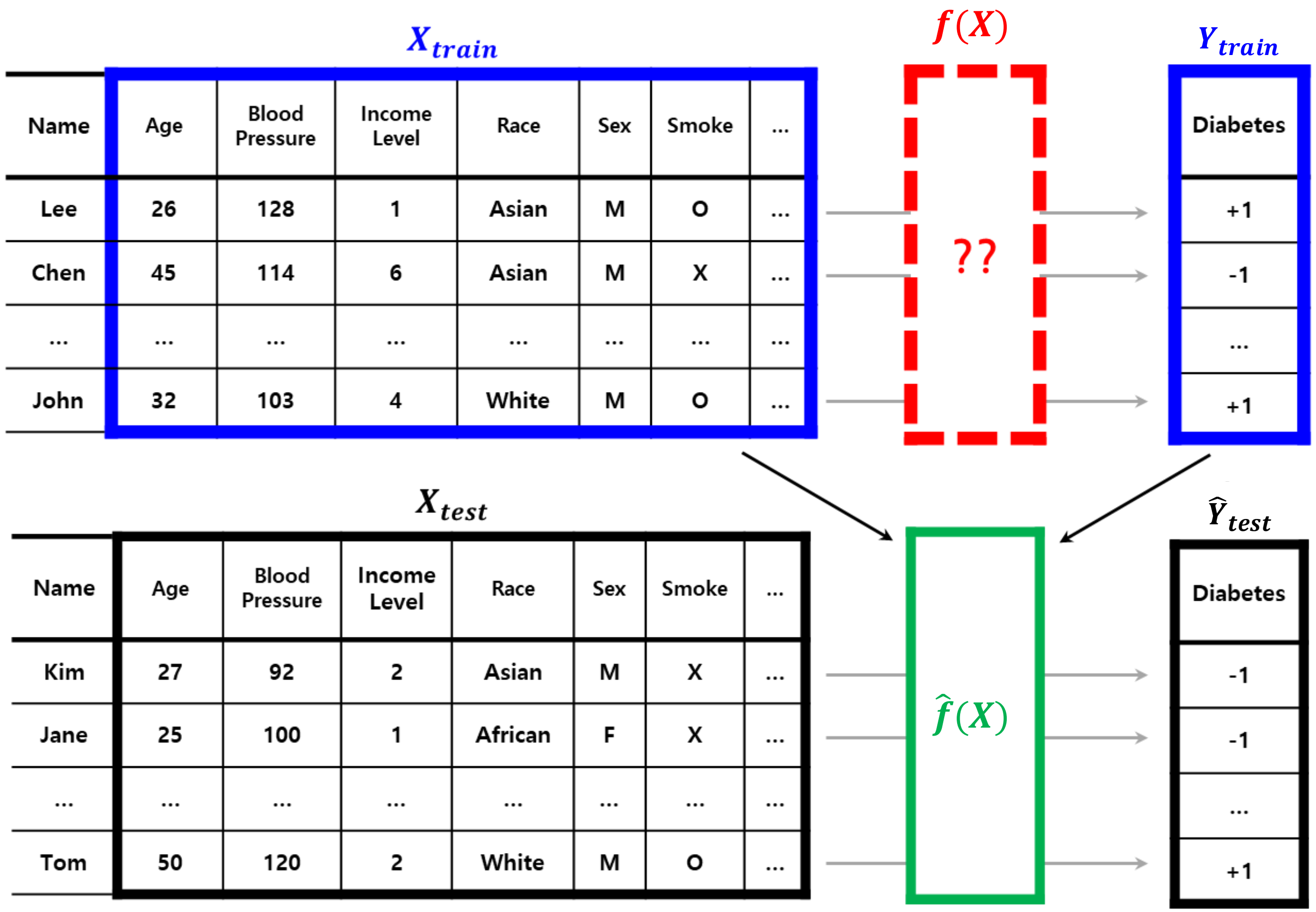
- Define a goal and problem type.
- Define data (X and Y).

- Data collection
- Data preprocessing
- Data splitting

- Learning
- Evaluation

- Test (Inference on new data)
- Monitor and feedback

Dataset and Model



Name	Age	Blood Pressure	Income Level	Race	Sex	Smoke	...
Lee	26	128	1	Asian	M	O	...
Chen	45	114	6	Asian	M	X	...
...	...	...	...	...	...	...	...
John	32	103	4	White	M	O	...

Diabetes
+1
-1
...
+1

Name	Age	Blood Pressure	Income Level	Race	Sex	Smoke	...
Kim	27	92	2	Asian	M	X	...
Jane	25	100	1	African	F	X	...
...	...	...	...	...	...	...	...
Tom	50	120	2	White	M	O	...

Diabetes
-1
-1
...
+1

An example of machine learning-based diabetes prediction model using patient information

Data Collection

"A data dictionary is a collection of names, definitions, and attributes about data elements."

Data

client_id	name	dob	gender	marital_status	current_address	description
1	Ki Ding	03/02/01	M	Single	Osaka, Japan	-
2	Gu Fing	30/08/99	M	Single	Tokyo, Japan	Certificate for proof of date of birth is yet to be submitted
3	Joe King	02/11/99	M	Married	Nagoya, Japan	-

Data dictionary (Metadata)

	Column	*Data type*	*Field size*	*Description*
1	client_id	int	5	Client's ID
2	name	nvarchar	30	Client's fullname
3	dob	date	8	Date of birth as per client's documents
4	gender	char	2	M – Male, F – Female, NB – Non-binary
5	marital_status	char	30	Marital Status as described by the client
6	current_address	char	300	Current residential address as described by client
7	description	nvarchar	300	Notes

Data Preprocessing

"Garbage in, Garbage out."

Your analysis is as good as your data.

1. **Data Cleaning**

   Improves data quality by removing errors in the data.

2. **Data Integration**

   Integrate different datasets according to the purpose of analysis.

3. **Data Transformation**

   Transforms data into a form that can be applied with analysis algorithms.

4. **Data Reduction**

   Intentionally reduces the data to increase the ease and efficiency of analysis while minimizing loss of information.

Data Preprocessing

"For unstructured data, data encoding is required for machine learning application."

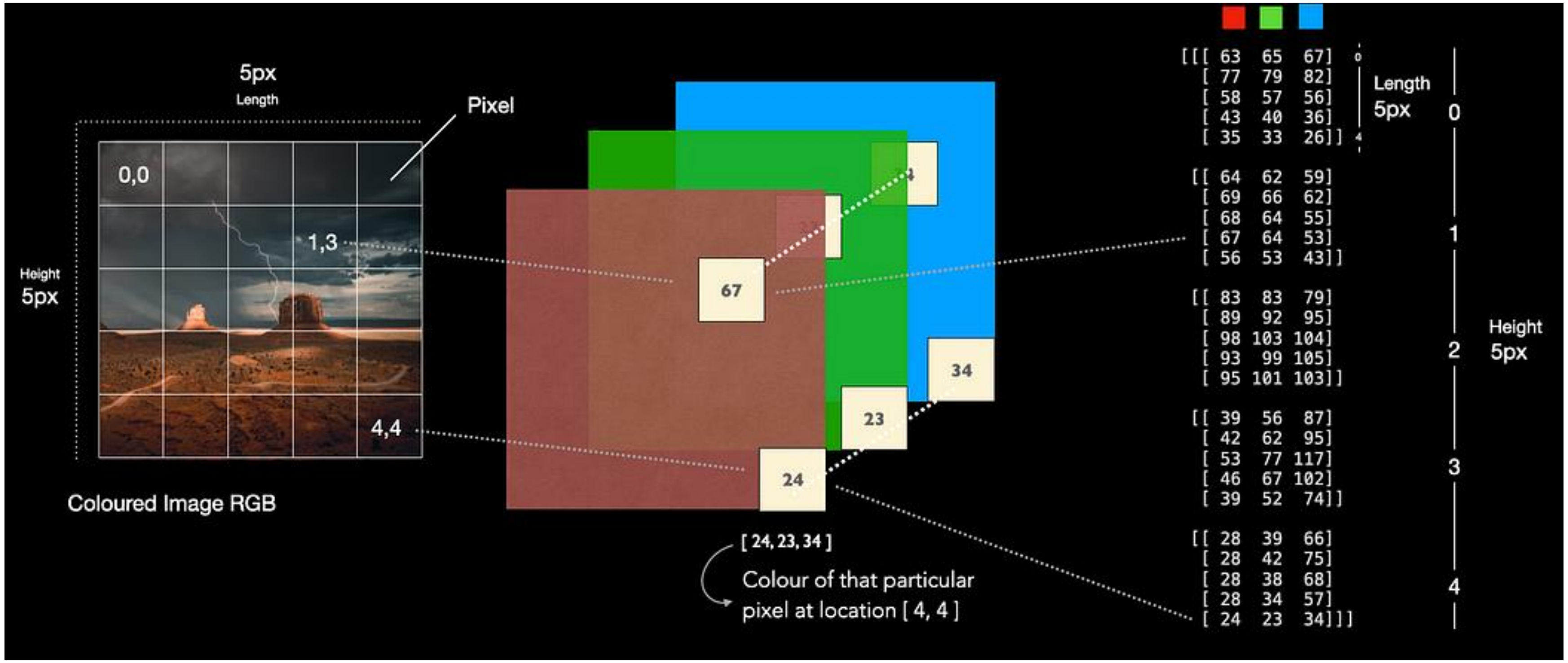
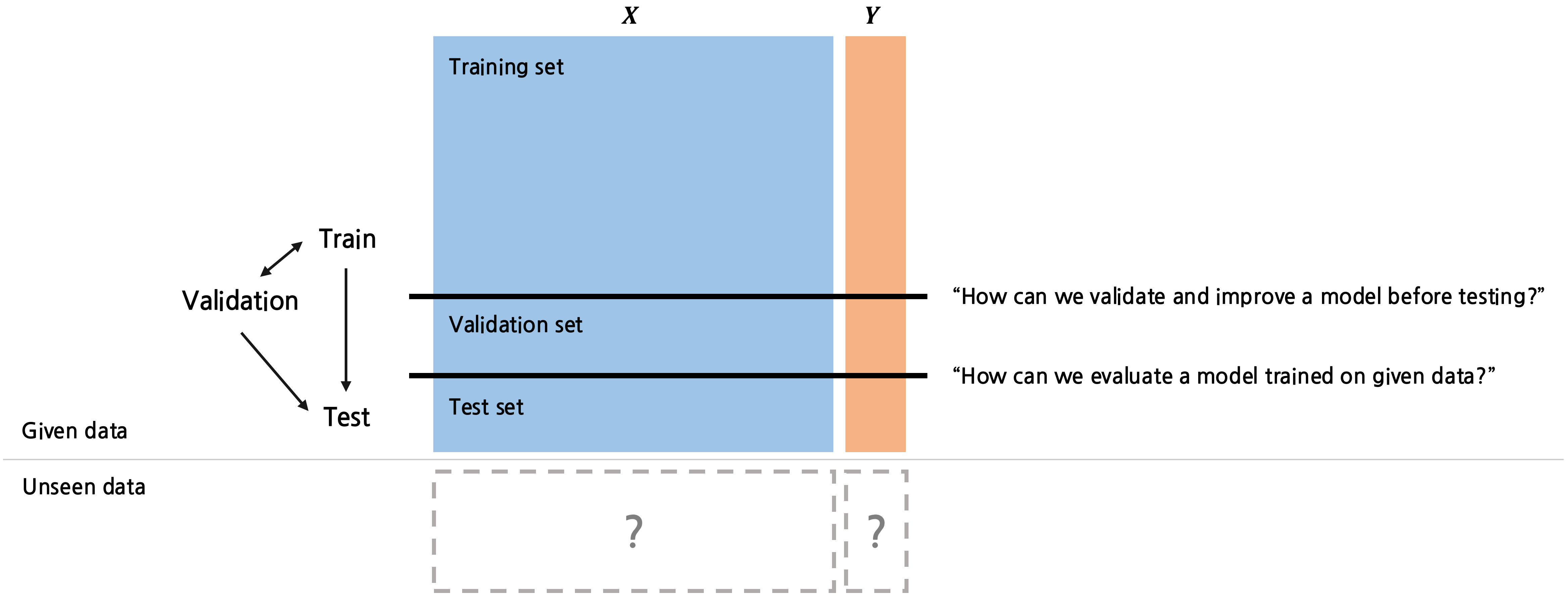


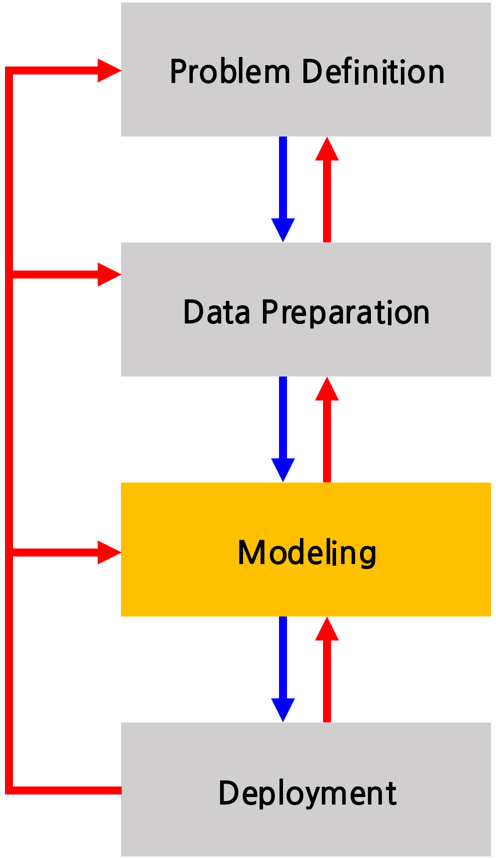
Image Source | Lehn, F. vom. (2023, August 19). Understanding the Structure of RGB Images and How Pixel Values Represent Color. *Advanced Deep Learning*. https://medium.com/advanced-deep-learning/decoding-image-representation-understanding-the-structure-of-rgb-images-6a211eb8800d

Data Splitting

“A given dataset is divided into training, validation, and test sets for model selection and evaluation.”



Machine Learning Pipeline



- Define a goal and problem type.
- Define data (X and Y).

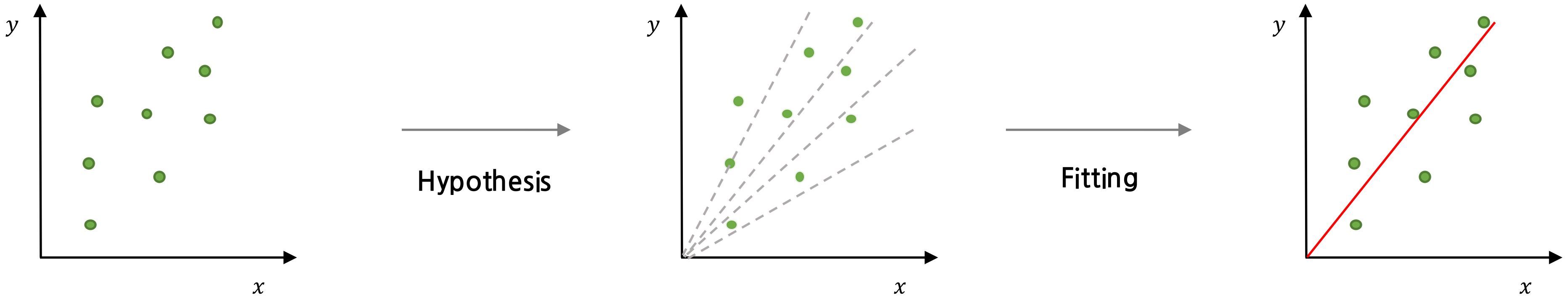
- Data collection
- Data preprocessing
- Data splitting

- Learning
- Evaluation

- Inference on new data
- Monitor and feedback

What is *Modeling*?

"Modeling is the process of finding a model that best describes the patterns of data."

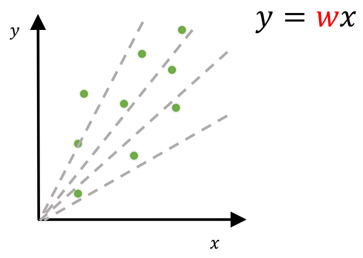


"In this data, x and y probably have a linear correlation."

$y = wx$

$y = 2x$

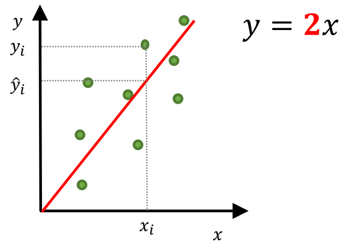
The 4 Steps for Machine Learning Modeling



1. "Define a model."
   - Which model to use for the given data and problem?

$$\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

2. "Define an objective function."
   - By what criteria should the optimal model be found?



3. "Fit the model."
   - Which model optimizes the objective function for the training data?

Learning

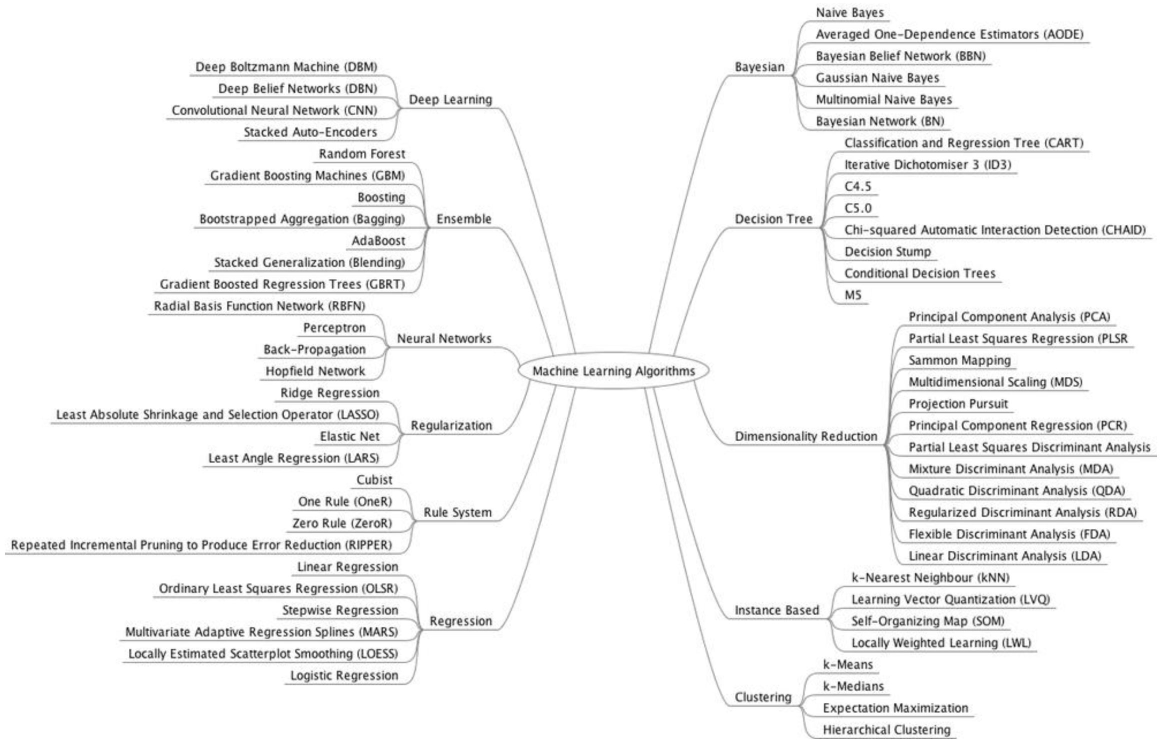
$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

4. "Evaluate the model."
   - What is the performance of the trained model?

Evaluation

The 4 Steps for Machine Learning Modeling: (1) "Define a model."

"Understand given data and problem types and assume patterns underlying in the data."



Different machine learning models

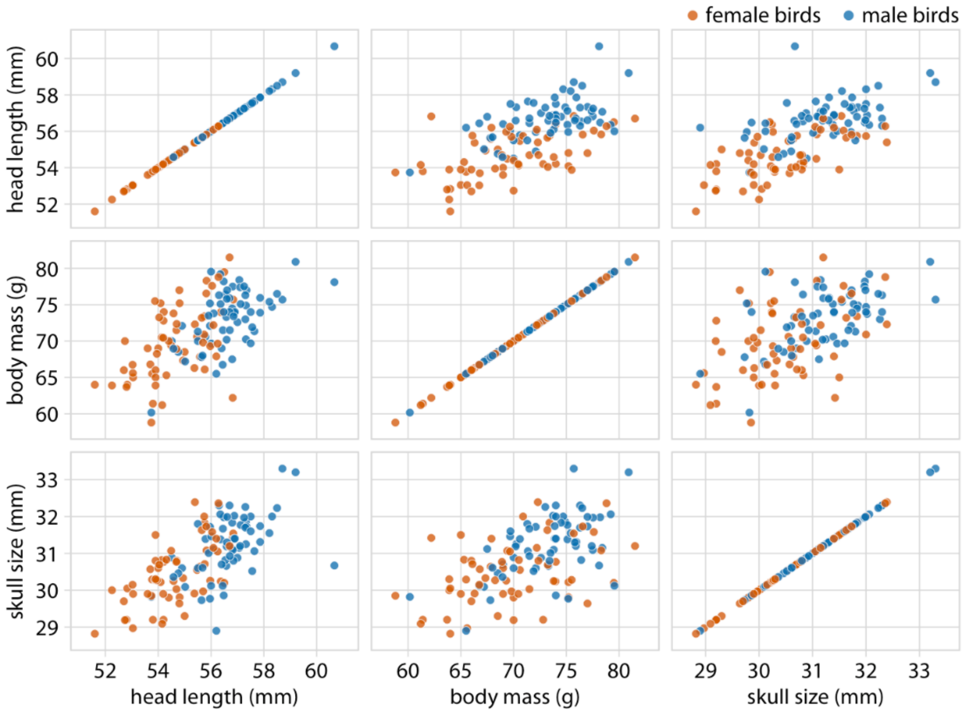
(Source | https://www.deepmarketer.com/blog/2017/1/30/machine-learning-algorithm-taxonomy)

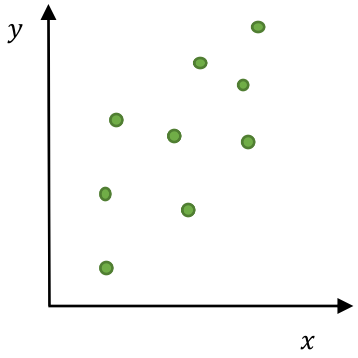
"Which of so many ML models to choose.."

- Understand the data through EDA.
- Find a model that fits the data and problem.
  (structured/unstructured, regression/classification, …)
- Make a hypothesis.
- There is no free lunch!
  (Trade-off by model complexity)

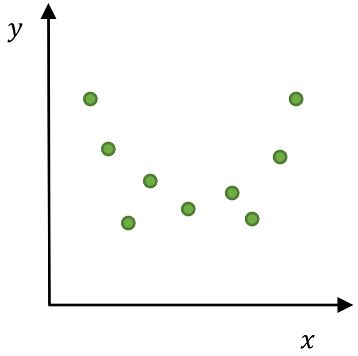
The 4 Steps for Machine Learning Modeling: (1) "Define a model."

"The answer is in the data."



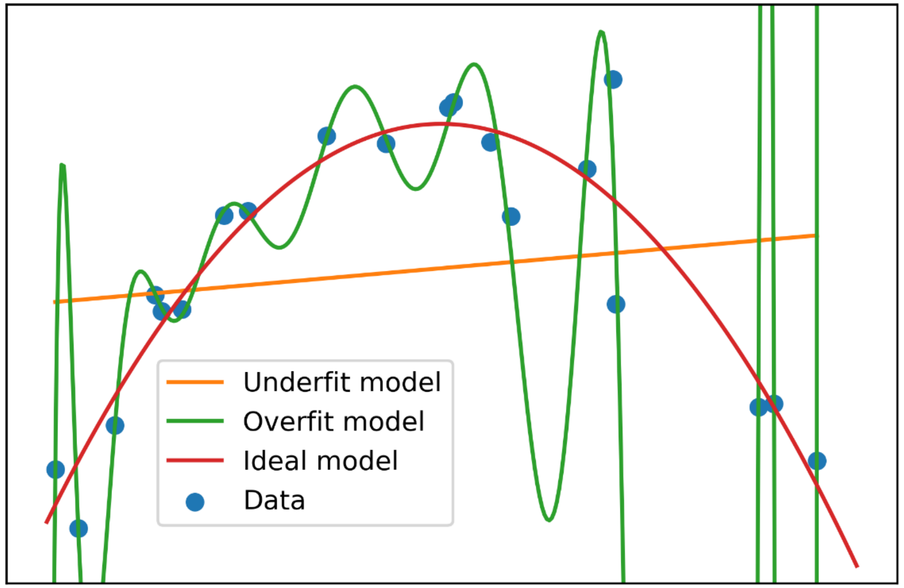


$$y = w_0 + w_1 x$$



$$y = w_0 + w_1 x + w_2 x^2$$

The 4 Steps for Machine Learning Modeling: (1) "Define a model."

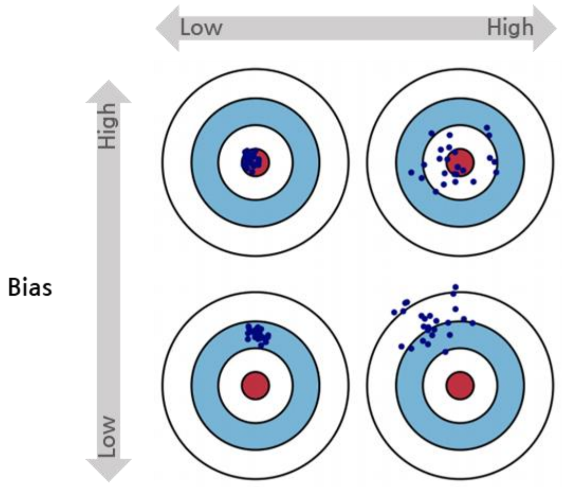


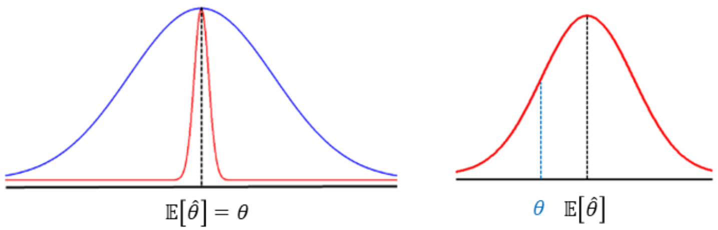
Example: a polynomial regression

- **Model**
  - $y = f(x; \boldsymbol{w}) = \sum_{j=0}^{M} w_j x^j$
  - Defined by the user's assumption.
- **Parameter**
  - $w = [w_0, w_1, \dots, w_M]$
  - Learnable by training data.
- **Hyperparameter**
  - M
  - User-defined parameters to control the learning process (*i.e.*, model and parameters).

Bias-Variance Tradeoff

"The bias-variance tradeoff describes the relationship between a model's complexity"





Unbiased

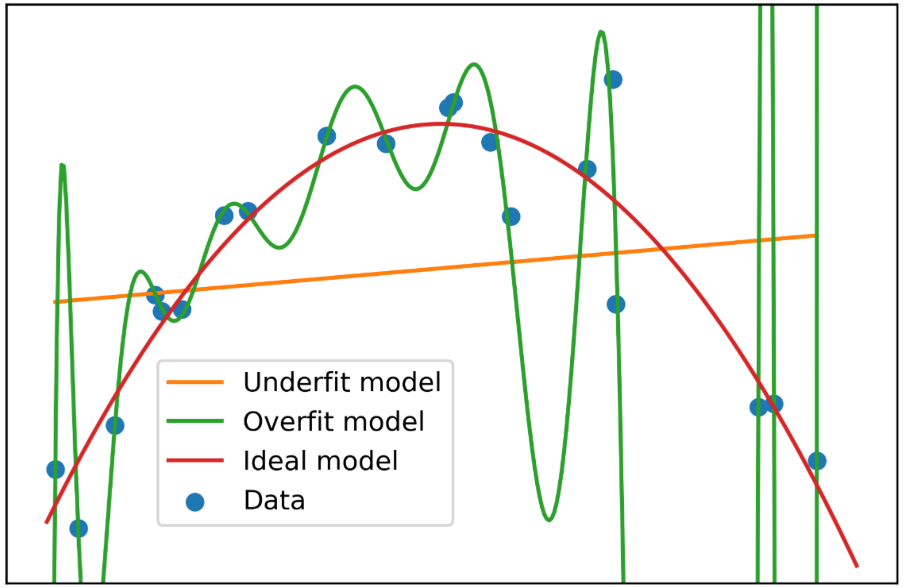
(small and large variance)

Biased

- Bias-variance decomposition of mean squared error
  - Training data: $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
  - Model: $y = f(x) + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$
  - $\mathbb{E}_{D,\epsilon}\left[\left(y - \hat{f}(x;D)\right)^2\right] = \left(\mathrm{Bias}_D\left[\hat{f}(x;D)\right]\right)^2 + \mathrm{Var}_D\left[\hat{f}(x;D)\right] + \sigma^2$
  - $\mathrm{Bias}_D\left[\hat{f}(x;D)\right] = \mathbb{E}_D\left[\hat{f}(x;D) - f(x)\right]$
  - $\mathrm{Var}_D\left[\hat{f}(x;D)\right] = \mathbb{E}_D\left[\left(\mathbb{E}_D\left[\hat{f}(x;D)\right] - \hat{f}(x;D)\right)^2\right]$

The 4 Steps for Machine Learning Modeling: (1) "Define a model."

"There are pros and cons depending on the complexity of the model."

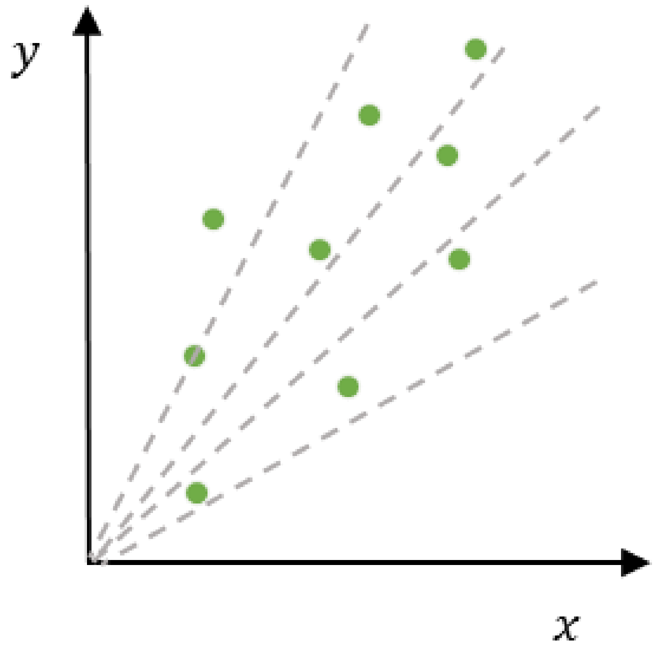


Overfitting and underfitting of a model

	Simple	Complex
Assumption	Strong	Weak
Interpretability	High	Low
Learning	Easy	Hard
Generalization	Underfit	Overfit

The 4 Steps for Machine Learning Modeling: (2) “Define an objective function.”

“An objective function mathematically formulates the goal of the problem and use it as a criterion for finding the optimal model.”



- $\hat{y} = f(x) = wx$
- What is the optimal value, w^*, for w?
- The cost function:

$$L(w) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

The objective function:

$$\min_{w} L(w)$$

- The optimal solution for w :

$$w^* = \arg\min_{w} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

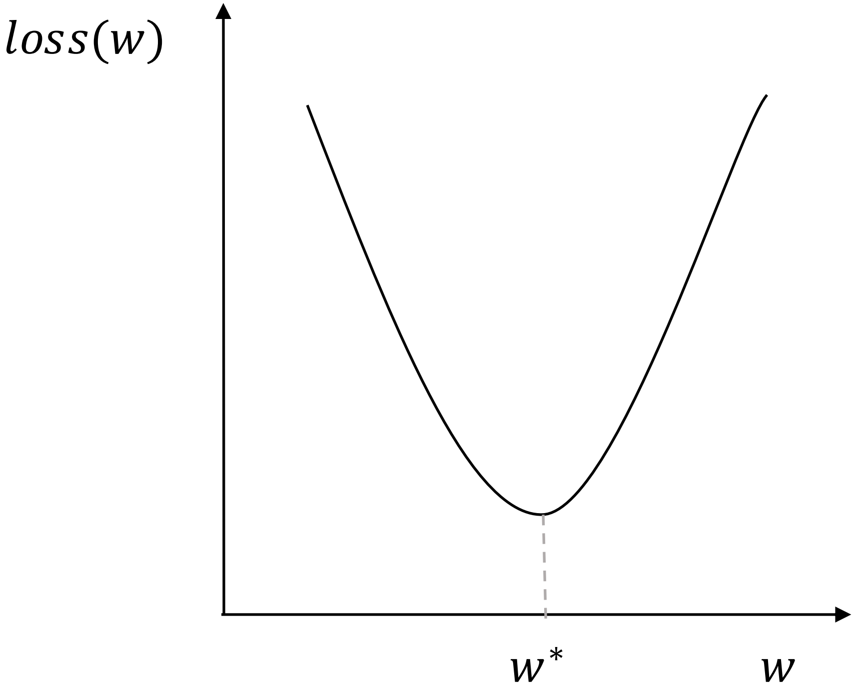
The 4 Steps for Machine Learning Modeling: (2) "Define an objective function."

"A loss function computes the distance between the current output of the algorithm and the expected output.
A cost function computes the average loss over the entire training dataset."

- **Regression Loss Functions**
  - Squared error loss (a.k.a. $L2$ loss): $L = (y - \hat{y})^2$
    - SSE (Sum of Squared Errors): $SSE = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
    - MSE (Mean Squared Errors): $MSE = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / N$
  - Absolute error loss (a.k.a. $L1$ loss): $L = |y - \hat{y}|$
    - MAE (Mean Absolute Errors): $L = \sum_{i=1}^{N} |y - \hat{y}| / N$
- **Classification Loss Functions**
  - Cross entropy loss: $L = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$ where $y \in \{0, 1\}$
  - Hinge loss: $L = \max(0, 1 - y\hat{y})$ where $y \in \{-1, 1\}$

The 4 Steps for Machine Learning Modeling: (3) "Fit the model."

"We find the values for the model parameters that minimize the cost function."



Gradient Descent Algorithm

1. Calculate the gradient ∇_t at w_t, which is the value for w at iteration t.

2. Update w_t with ∇_t and a learning rate η:

$$w_{t+1} = w_t - \eta \nabla_{\mathrm{t}}$$

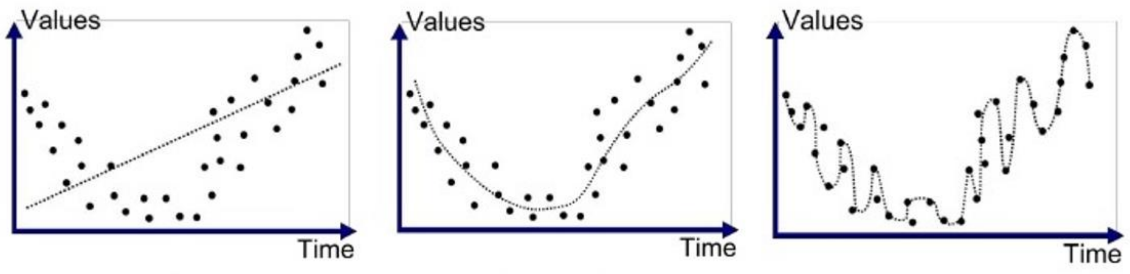
3. Calculate ∇_{t+1}.

   Stop if the stopping criteria are met. (i.e., $\nabla_{t+1} < \delta$)

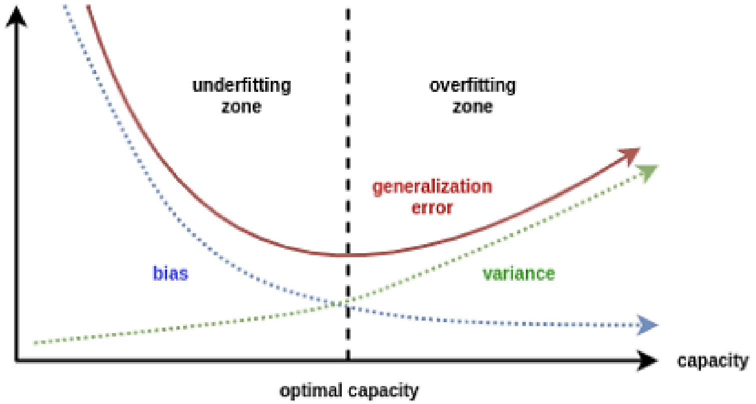
   Otherwise, repeat step 1.

The 4 Steps for Machine Learning Modeling: (4) “Evaluate the model.”

“Evaluate the generalization error with a validation dataset that was not used for training.”



Underfitted Good Fit/Robust Overfitted

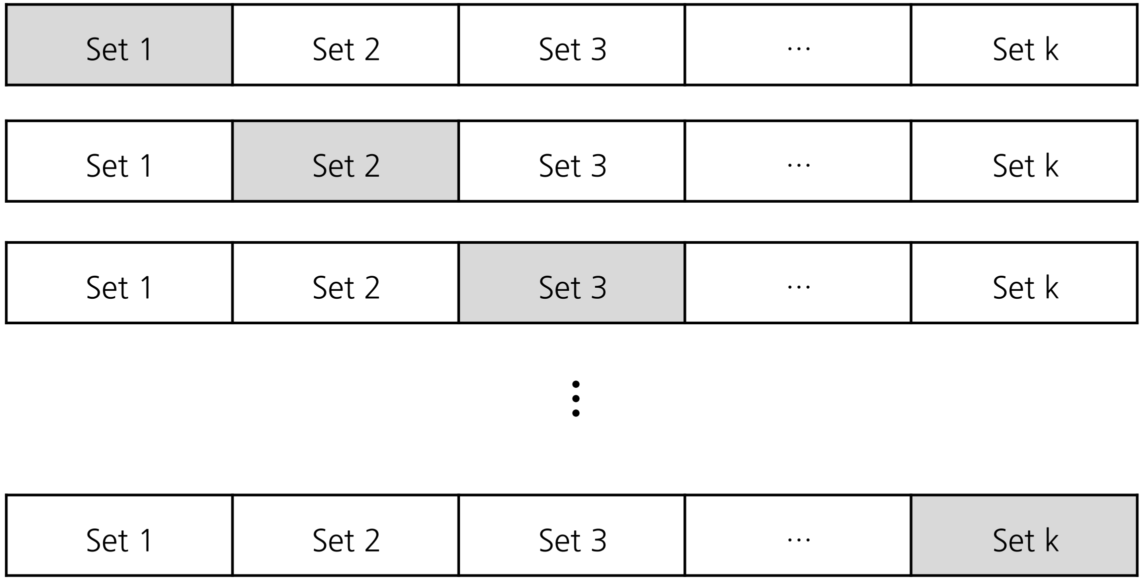


Overfitting and bias-variance trade-off

Image Source | https://medium.com/@cs.sabaribalaji/overfitting-6c1cd9af589

The 4 Steps for Machine Learning Modeling: (4) "Evaluate the model."

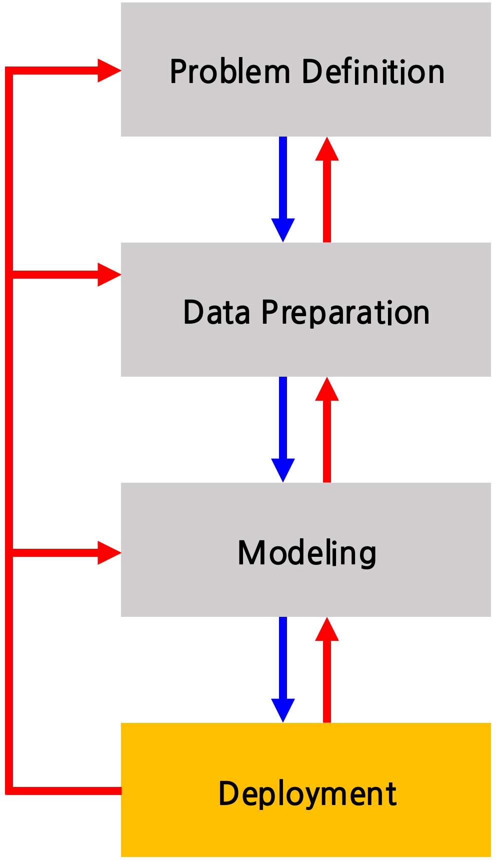
"We can use cross validation to evaluate the generalization performance of the model."



k-fold cross-validation

1. Divide the training dataset into k subgroups.
2. Except for 'Set 1', train the model on the rest.
   'Set 1' is then used to evaluate the performance of the trained model.
3. Repeat step 2 for the k subsets.
4. Compute the average of k performance measures.

Machine Learning Pipeline



- Define a goal and problem type.
- Define data (X and Y).

- Data collection
- Data preprocessing
- Data splitting

- Learning
- Validation (model selection, hyperparameter tuning)

- Test (Inference on new data)
- Monitor and feedback

MLOps

"The operation of machine learning models requires many factors."

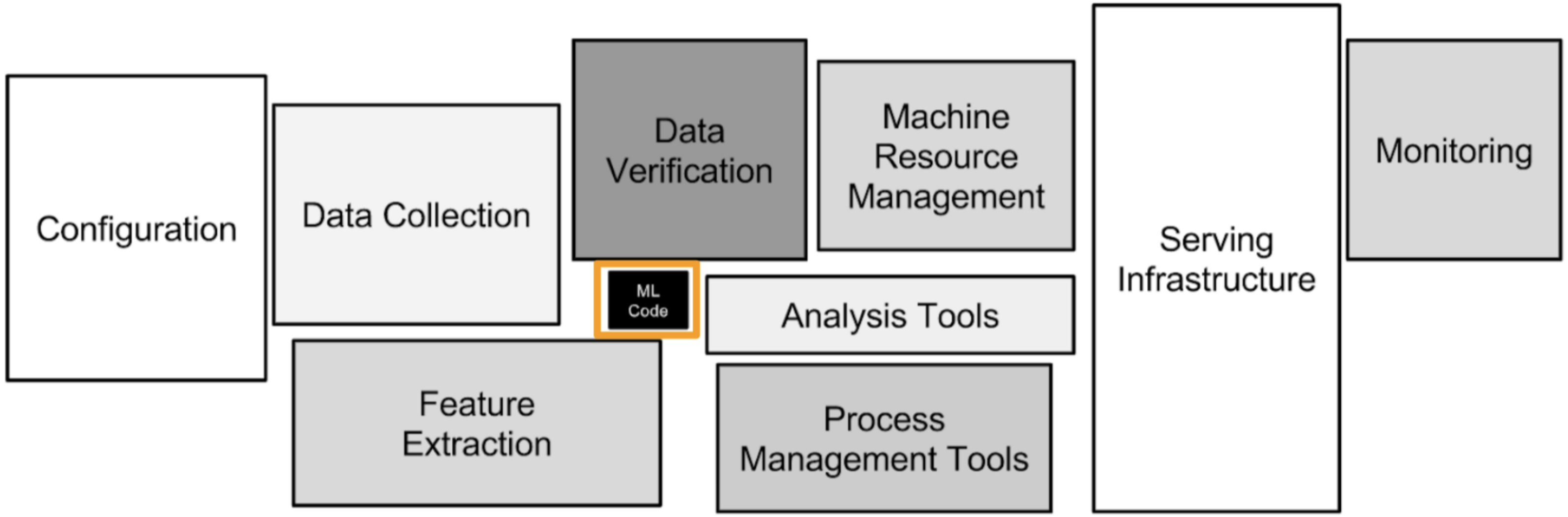
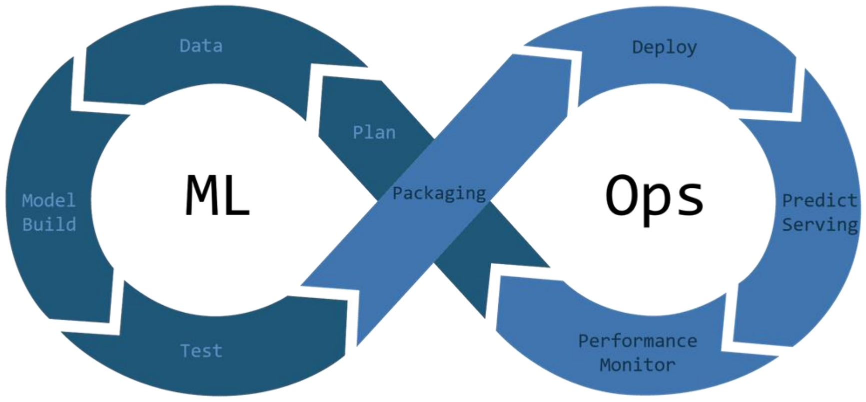


Image Source | Sculley, David, et al. "Hidden technical debt in machine learning systems." Advances in neural information processing systems 28 (2015): 2503-2511

MLOps

ML (Machine Learning) + Ops (Operations) = MLOps



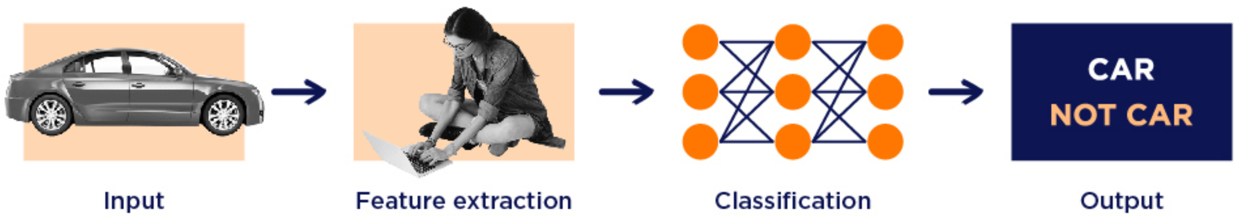
<u>Purpose of MLOps</u>

- Easy model development
- Stable and efficient operation of ML models
- Minimization of human error with automation
- ML model quality management

Definition of Deep Learning

Definition of Deep Learning

"Deep learning is the subset of machine learning that uses deep neural networks with representation learning."



Traditional Machine Learning before Deep Learning

Deep Learning

Image Source | www.iteratorshq.com (the original image is edited by the author)

Representation Learning

"How can we have good features for the given task?"

"What are the features of airplanes
that distinguish them from vehicles?"

"Wings?"

"Long body?"

Deep Learning as Representation Learning

"Deep learning extracts features from the training data by itself."

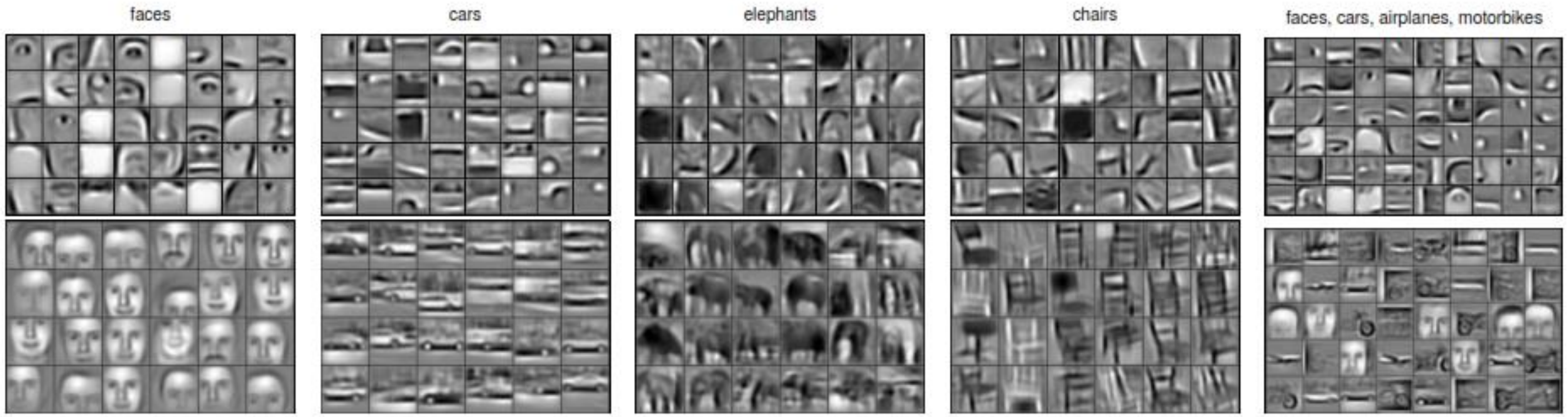
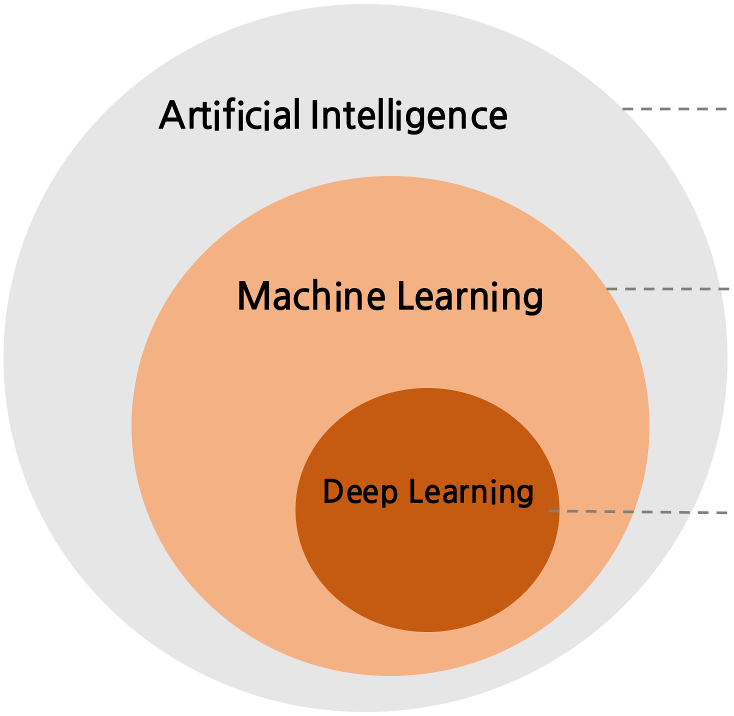


Figure 3. Columns 1-4: the second layer bases (top) and the third layer bases (bottom) learned from specific object categories. Column 5: the second layer bases (top) and the third layer bases (bottom) learned from a mixture of four object categories (faces, cars, airplanes, motorbikes).

Image Source | Lee, Honglak, et al. "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations." *Proceedings of the 26th annual international conference on machine learning*. 2009.

AI, Machine Learning, and Deep Learning



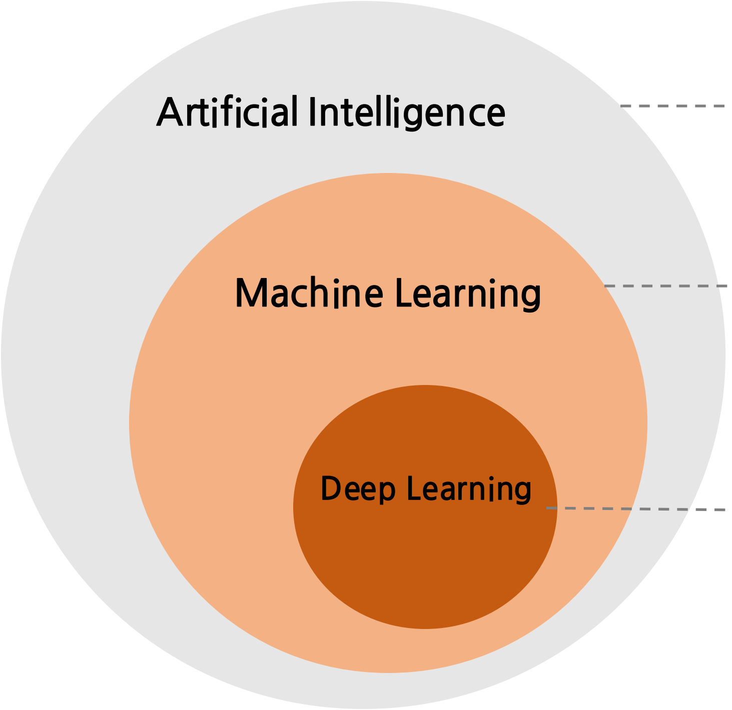
Any technique that enables computers to mimic intelligence

AI techniques that enable computers to automatically learn from data to produce a required output for their user

A subset of machine learning that deals with deep neural networks

Takeaways

AI, Machine Learning, and Deep Learning



Any technique that enables computers to mimic intelligence

AI techniques that enable computers to automatically learn from data to produce a required output for their user

A subset of machine learning that deals with deep neural networks

Thank you! 😀