1. Introduction to Machine Learning

KUBIG 학술부

Index

KU-BIG

Example) Which image contains the image of tree?

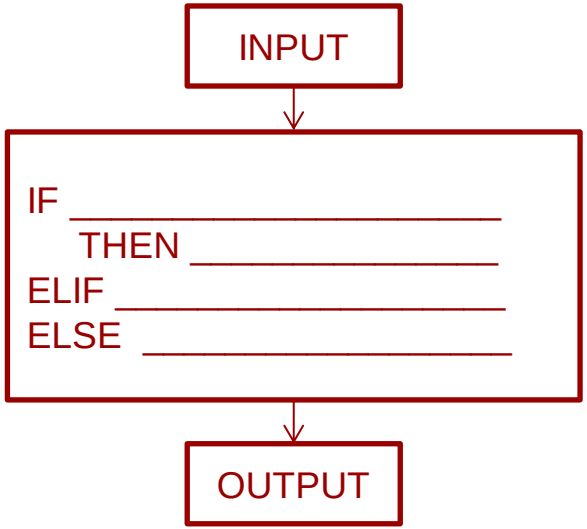
1. What is Machine Learning ?

" Field of study that gives computers the ability to learn without being explicitly programmed " - Arthur Samuel (1959)

"A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.“

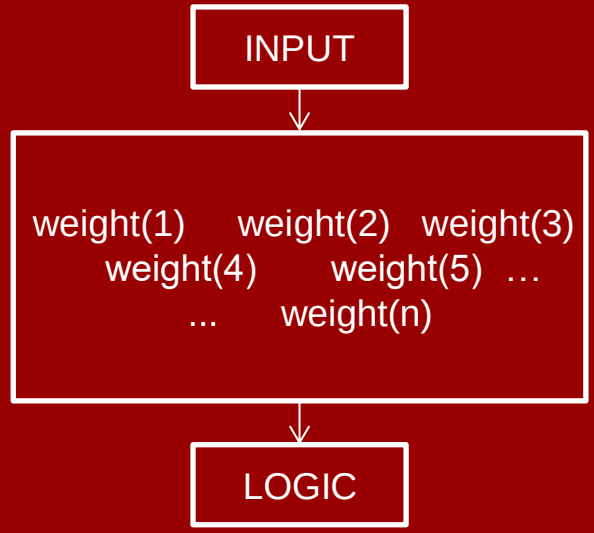
- Tom. M. Mitchell (1997)

Explicit Programming



Line-by-line coding with specific rule

Data Based Programming



Converting the conditions to specific weights for the learning model

KU-BIG

Example) Google “the Manhattan Project

- **인공지능 맨해튼 프로젝트 (AI Manhattan Project)**
  - 인공지능 기술 개발을 위한 구글의 세기적 프로젝트
  - 관련 IT 기업과 전문가들을 확보하기 위한 거대 인수합병을 진행
- 1000명 이상의 인력이 구글의 인공지능/머신러닝 연구 진행 중 (아주경제)
- 2014년 1월 Deep Mind 인수 (6억 5000만 달러)
- 2014년 1월 Nest Labs 인수 (32억 달러)
- 기타 50여 개의 인공지능 기업들이 맨해튼 프로젝트에 참여, 연구 진행

The Manhattan Project

- During the World War 2, the project was to make the world’s first nuclear weapon ahead of Hitler’s Nazi
- Many renowned scientists gathered officially funded by the US government

Data Revolution plan by Google

- The project to stay ahead of the advanced AI technology
- Deepmind AI reduces Google Data Center cooling bill by 40%

2. Basic Idea behind Machine Learning

[The training data set]

X	Y
{3,6,9}	3
{2,5,7}	2
{2,3,5}	1

training

[The ML model]

Q. X = {4,7,8}

A. Y = ????

KU-BIG

Example) Self-Driving Car

[google image) Self-Driving Car in Las Vegas provided by Lyft]

Knowledge-Driven Programming

- Consider every possibility with human knowledge
- 100% controlled by pre-coded algorithm
- Concluded Infeasible(Too many Rules)

Data-Driven Programming

- Use logic driven from the computer based on the given data
- Reinforcement Learning
- Commercialized in Las Vegas

3. What are the kinds of Machine Learning ?

Supervised Learning

- Regression
- Classification

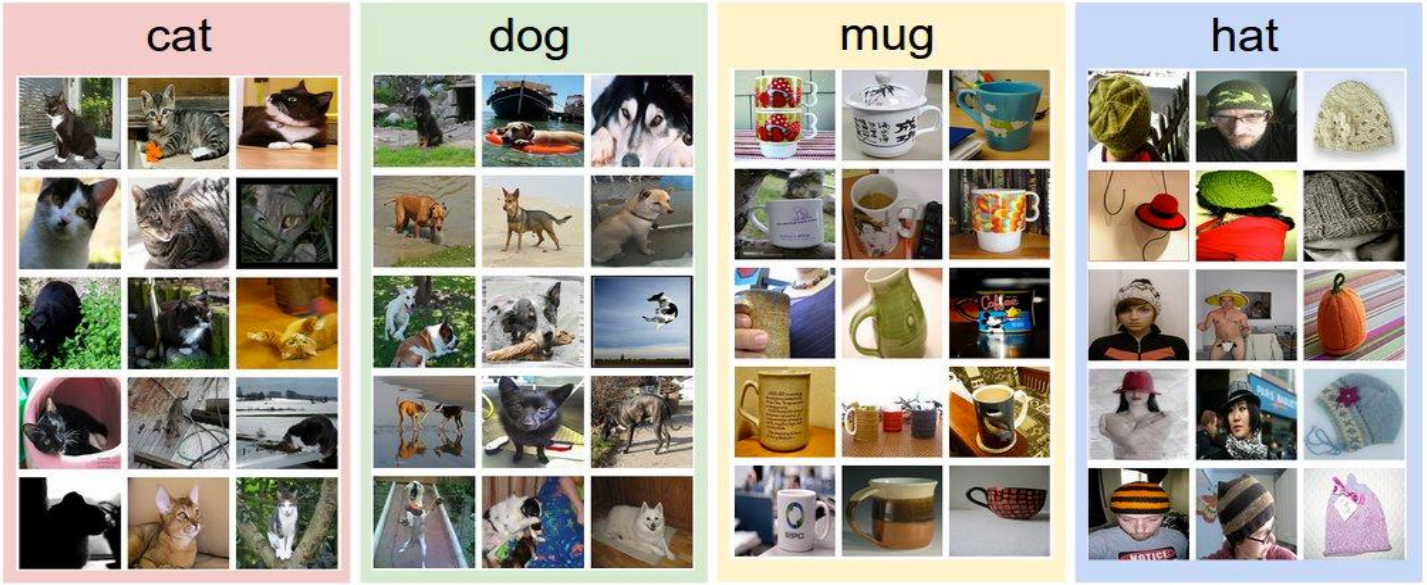
Semi- Supervised Learning

Unsupervised Learning

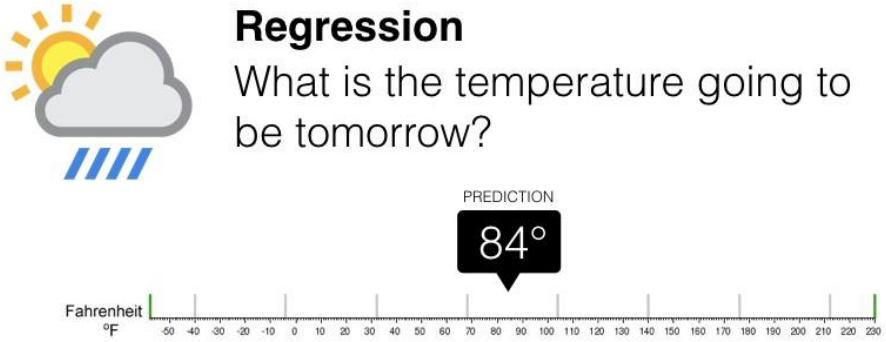
- Clustering
- Visualization
- Dimensionality Reduction
- Association Rule

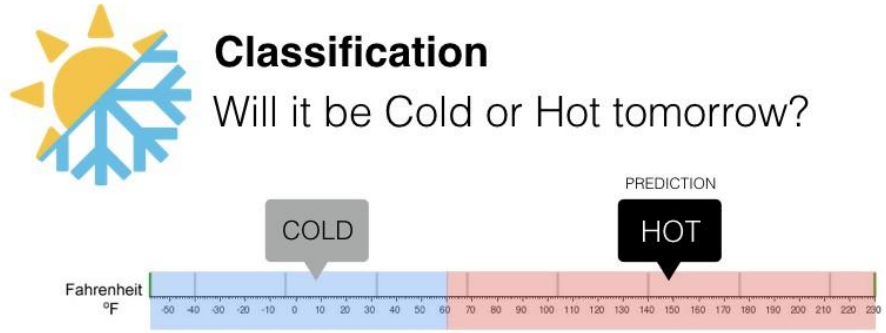
Reinforcement Learning

3-1. Supervised Learning



The Label(pre-determined output value) is given : Data in a form of (X, y)





Regression

- Output variable is continuous
- Predict y(the target variable) with given X(the predictor variable)
- Ex) temperature, value, price etc.

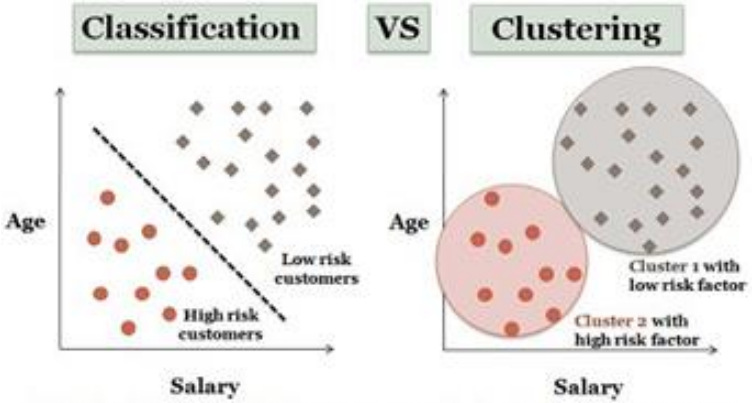
Classification

- Output Variable is discrete
- Binary Classification [1 or 0]
- Multi-label Classification [a \ b \ c \ d]

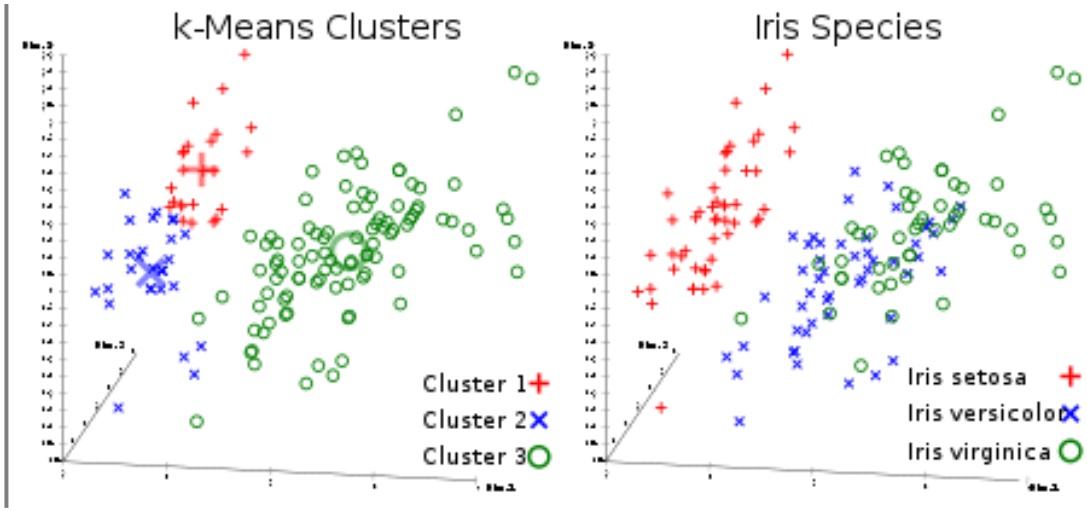
Example of Supervised Learning

- KNN (K^{th} Nearest Neighbors)
- Linear Regression
- Logistic Regression
- SVM (Support Vector Machine)
- Decision Tree
- NN (Neural Net)

3-2. Unsupervised Learning



Risk classification for the loan payees on the basis of customer salary



- The Label(pre-determined output value) is not given : Data in a form of (X)

Supervised Learning

X	y
10	90
9	80
3	50
2	30

Both X(predictor variable) and y(targe variable) are given

Unsupervised Learning

X
10
9
3
2

NO LABEL ; no y(target variable) given

Example of Unsupervised Learning

Clustering

- K-Means Clustering
- Hierarchical Clustering
- DBSCAN (Data-Based Spatial Clustering of Applications with Noise)

Visualization / Dimensionality Reduction

- t-SNE (t-distribution Stochastic Neighbor Embedding)
- PCA (Principal Component Anlaysis)
- Kernel PCA

The Association Rule

Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \,\&\, C \Rightarrow D$	1/5	1/3	5/9

지지도(support) : $\Pr(A \cap B) = \frac{A와\ B\ 둘다\ 고른\ 사람\ 수}{전체\ 분석하는\ 총\ 인원\ 수}$

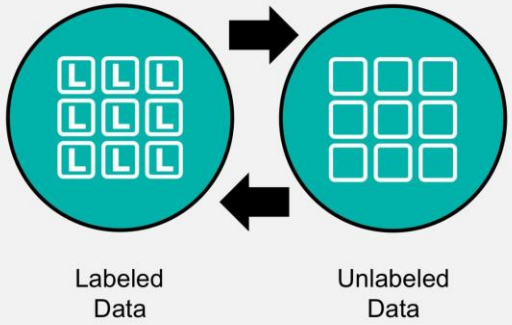
신뢰도(confidence) : $\Pr(A \mid B) = \frac{A와\ B\ 둘다\ 고른\ 사람\ 수}{B를\ 고른\ 사람\ 수}$

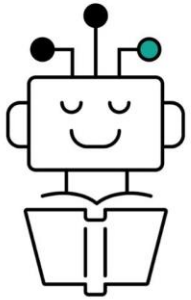
향상도(lift) $\frac{Pr(A \mid B)}{Pr(B)} = \frac{Pr(A \cap B)}{Pr(B) \star Pr(B)}$

KU-BIG

3-3. Semi-Supervised Learning

Semi-supervised learning uses a combination of supervised labeled training data with unsupervised methods:



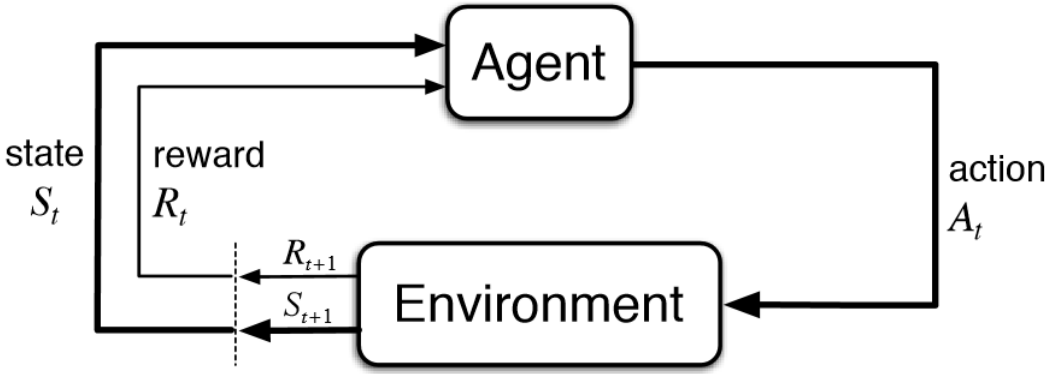


Semi-Supervised Learning

1. Load labeled input training data.
2. Model is trained on the data.
3. Present unlabeled raw data.
4. Algorithm infers classifiers for unlabeled data on its own.
5. High-confidence data is added to labeled training data set.
6. Algorithm progressively adapts and learns.

▷ The Label(pre-determined output value) is partially given

3-4. Reinforcement Learning



Each Agent updates the Policy the Action according to the observed Environment in a way that maximizes the Reward and minimizes the Penalty

Example) Google Deepmind Challenge [Alpha Go]

[google image) Sedol Lee the only person to defeat Alpha Go]

Reinforcement Learning

- Database of every professional players
- Use Reinforcement Learning method for the training
- Every decision made for the optimal outcome

New generation : AlphaGo [zero]

- No input of human data anymore
- Endless re-generation of the random training data
- Strategies that was never seen before

KU-BIG

ML visualized map