



Chapter 1

AI Background

인공지능과 머신러닝

| 인공지능 (Artificial Intelligence) 이란

인공지능 (Artificial Intelligence) 의 사전적 의미

- '인간의 지능으로 할 수 있는 사고, 학습, 자기 개발 등을 컴퓨터가 할 수 있도록 하는 방법을 연구하는 컴퓨터 공학 및 정보기술의 한 분야로서, 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것'
- 딥러닝 (Deep Learning)이 부흥하기 시작하면서 인공지능을 이야기하면 딥러닝을 바로 떠오르는 사람이 많지만, 사실은 전통적인 머신 러닝 (데이터 사이언스 또는 데이터 분석을 위한)기법을 포함하는 단어
- 인공지능을 쉽게 정의한다면 컴퓨터가 데이터를 가지고 학습할 수 있도록 하는 기술이라고 할 수 있음
(머신 러닝, 딥러닝 모두 포함)

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - Machine Learning

- “무엇(X)으로 무엇(Y)을 예측하고 싶다”

정형 데이터(행렬) - (Tabular data)

X						Y
id	레벨	경험치	게임시간	파티수	아이템정보	게임이탈여부
...

X: 게임 유저들의 게임 내 활동 정보, Y: 게임 이탈 여부 / 어뷰징 여부 -> 이상 탐지 모델

X: 고객들의 개인 정보 및 금융 관련 정보, Y: 대출 연체 여부 -> 대출 연체자 예측 탐지 모델, 대출 연체 관련 주요 feature 추출

X: 고객들의 상품 구매 내역 -> 매장내 상품 진열 위치 리뉴얼을 통한 매출 증대

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 (이미지처리)

- 이미지 분류

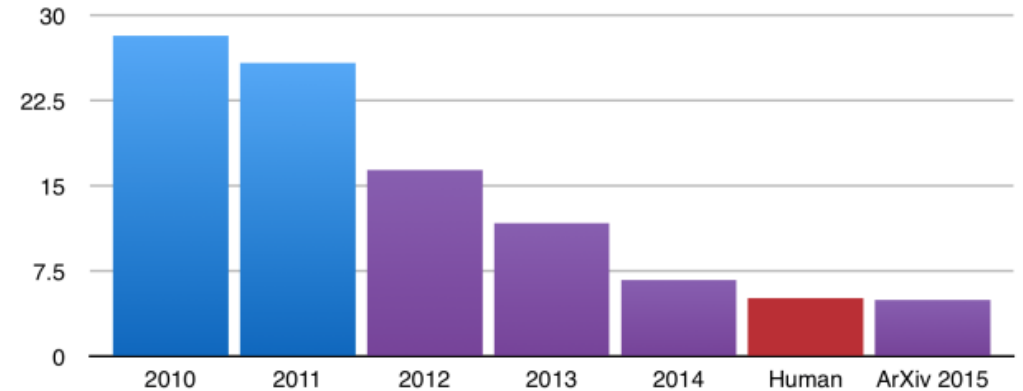


VS



고양이와 강아지를 분류하는 모델

※ 이미지/텍스트 - 비정형 데이터



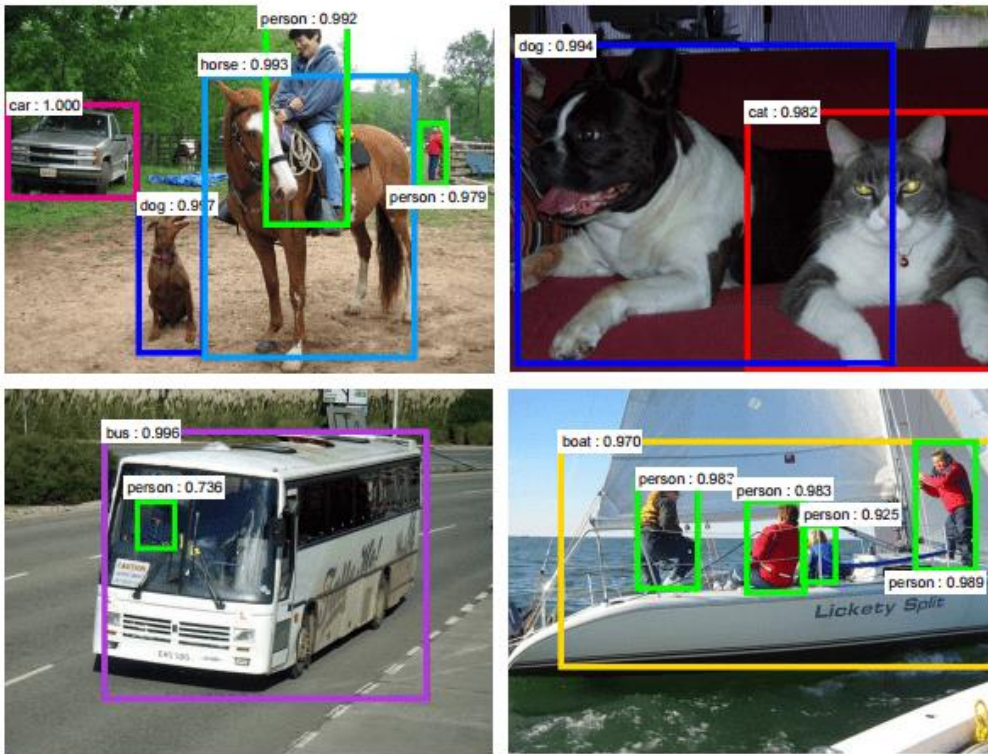
년도 별 ImageNet데이터 분류 모델 성능

- 인간이 이미지를 분류하는 성능은 약 95%
- 2015년에 ResNet이라는 모델이 약96%의 성능을 기록하면서 나오면서 인간의 성능을 뛰어넘기 시작
- 단순히 이미지를 분류하는 것을 넘어서 다양한 분야, 다양한 방식으로 발전
 - 적은 데이터로 높은 성능의 모델을 만들 수 있는 방법 / 더 강건한 (Robust) 모델을 만들 수 있는 방법 / Noise 데이터를 걸러 낼 수 있는 방법 등 다양한 형태와 분야로 발전

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 (이미지처리)

- 객체 탐지 (Object Detection)



Object detection이 적용된 이미지

- 이미지 및 비디오 속에 포함 되어있는 물체들에 대해서 해당 물체가 어떤 물체인지 분류하는 문제와 물체의 위치를 찾아내는 일
- 객체 탐지 기술은 자율주행자동차, CCTV 등 카메라 기술을 바탕으로 이루어지고 있는 제품 및 서비스에 최근 도입되고 있음

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 (이미지처리)

- Image Resolution



- 저화질 이미지를 고화질 이미지로 해상도를 복원

- Colorization



- 흑백 이미지를 color이미지로 복원

이미지출처 :

Kasem, H. M., Hung, K. W., & Jiang, J. (2019). Spatial Transformer Generative Adversarial Network for Robust Image Super-Resolution. *IEEE Access*, 7, 182993-183009.
Zhang, R., Isola, P., & Efros, A. A. (2016, October). Colorful image colorization. In *European conference on computer vision* (pp. 649-666). Springer, Cham.

| 인공지능 (Artificial Intelligence) 이란

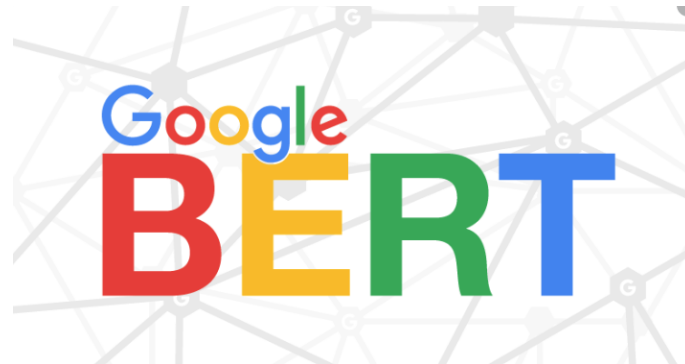
AI로 할 수 있는 것 - 자연어처리(Natural Language Processing : NLP)

- Sentiment Analysis (감정 분석)
 - 이 집은 맛집이야, '이 영화는 재미 없어' 이러한 문장 속 의미가 긍정인지 부정인지를 판단하는 것
- Summarization (요약)
 - 요약 분야는 주어진 Text에서 중요한 부분을 찾아내는 Extractive Summarization과 모델이 주어진 Text의 의미를 완전히 이해하여 이를 요약하는 새로운 문장을 만들어내는 Abstractive Summarization으로 나뉘어 짐

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 자연어처리(Natural Language Processing : NLP)

- Machine Translation (기계 번역)
 - 구글번역기, 파파고와 같이 다른 언어 간에 번역을 해주는 일
- Question Answering (질문 응답)
 - 주어진 문서를 이해하고, 문서 속 정보에 대한 질문을 했을 때 답을 이끌어내는 Task
- 2019년 즈음 BERT의 등장으로 NLP 분야는 급성장



| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 강화학습(Reinforcement Learning : RL)

- 알파고 (AlphaGo)
 - 알파고는 구글 의 딥마인드 (DeepMind)가 개발한 인공지능 바둑기사로 2016년 한국의 이세돌 기사와 대국해 4승 1패로 승리하며 세상을 놀라게 함
 - 알파고의 기본 원리는 강화학습으로 현재 상태 (바둑판)에서 어떠한 행동 (수)을 취해야 먼 미래에 보상이 최대 (승리)가 될 것인지에 대해서 학습하는 알고리즘
 - 알파고의 등장을 계기로 강화학습과 딥러닝을 결합한 심층 강화학습 (Deep Reinforcement Learning)의 연구가 활발히 진행되기 시작



| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Generative Adversarial Networks (GAN)
- 데이터를 '예측'하는 걸 넘어서 데이터를 직접 '생성' 해내는 모델이 등장
- GAN의 등장은 알파고의 등장과 비슷하게 인공지능 연구의 새로운 패러다임을 제시
- 최근에 연구되고있는 GAN의 성능은 이미 인간의 눈으로 구분하지 못할 정도의 고품질의 이미지 (또는 텍스트)를 생성해냅니다



BigGAN의 이미지 생성 예제

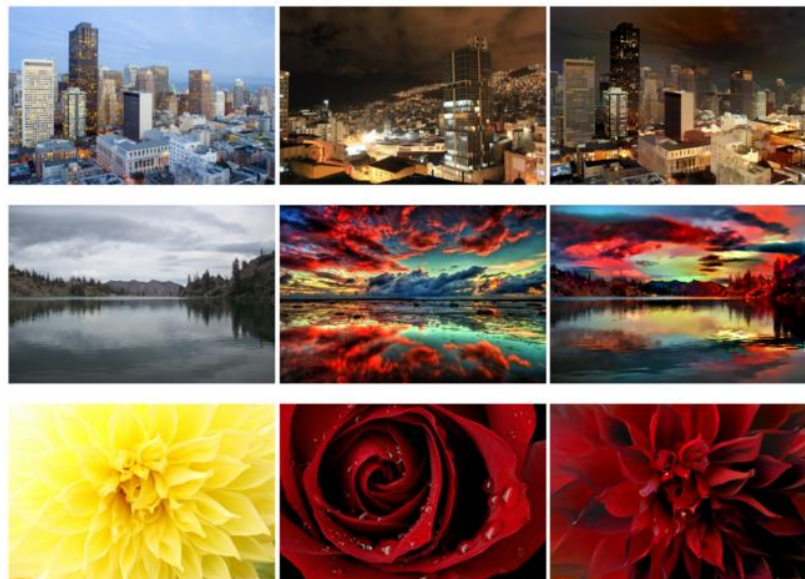
이미지출처 : Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096.

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Style Transfer

- GAN의 학습 방식을 활용하여 풍경의 사진을 밤 풍경의 사진으로 바꿔 준다거나, 여름 풍경을 겨울 풍경으로 style 을 transfer시킴

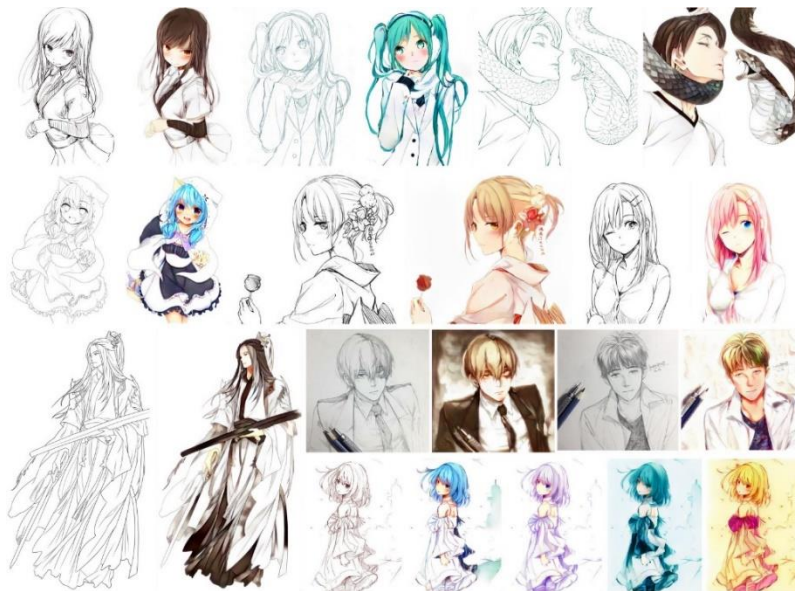


Style Transfer의 예제

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Style Transfer
- Style Transfer의 특징을 이용하여 아래 그림처럼 연필이나 펜으로 그린 그림에 대해 자동으로 채색해주는 GAN모형



Style Transfer for Anime Sketches 의 결과물

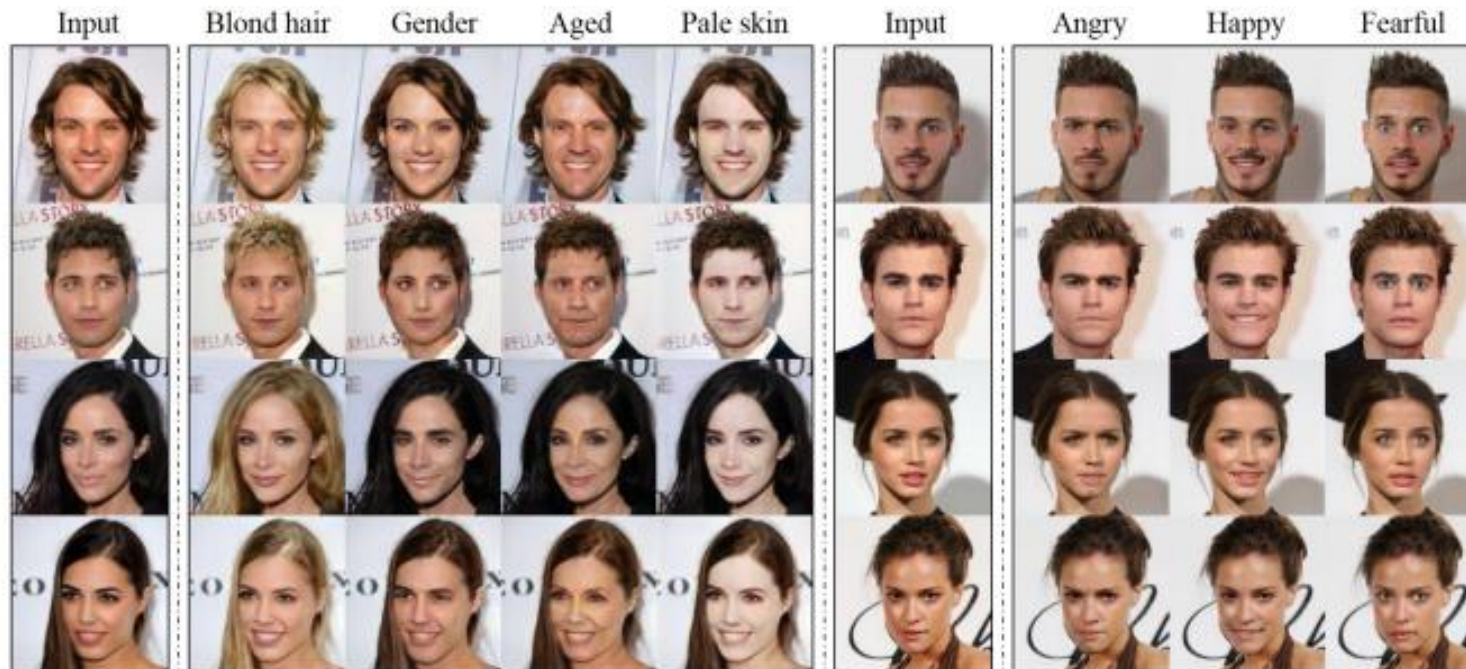
이미지출처 : Zhang, L., Ji, Y., Lin, X., & Liu, C. (2017, November). Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier gan. In 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR) (pp. 506-511). IEEE.

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- StarGAN

- Input 이미지에 대하여 참조 이미지의 피부 표정 등을 Transfer시키는 GAN 모형



StarGAN 의 결과물

이미지출처 : Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Deep fake
 - GAN의 등장은 합성 분야에서도 엄청난 발전을 이끌어냄
 - 아래 영상은 오바마 전 미국 대통령이 '트럼프는 쓰레기'라고 말하는 가짜(합성) 영상

<https://www.youtube.com/watch?v=cQ54GDm1eL0&feature=youtu.be>



Deep fake 기술이 적용된 오바마 영상

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Image completion



- 이미지의 빈 곳을 채워 주는 모델

Image completion이 적용된 예

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

- Everybody Dance Now
 - 다른 사람의 움직임을 포착해서 나에게 적용(영상 합성 기술)

<https://www.youtube.com/watch?v=PCBTZh41Ris&feature=youtu.be>



Everybody Dance Now 데모 영상

| 인공지능 (Artificial Intelligence) 이란

AI로 할 수 있는 것 - 분류를 넘어서 데이터를 생성

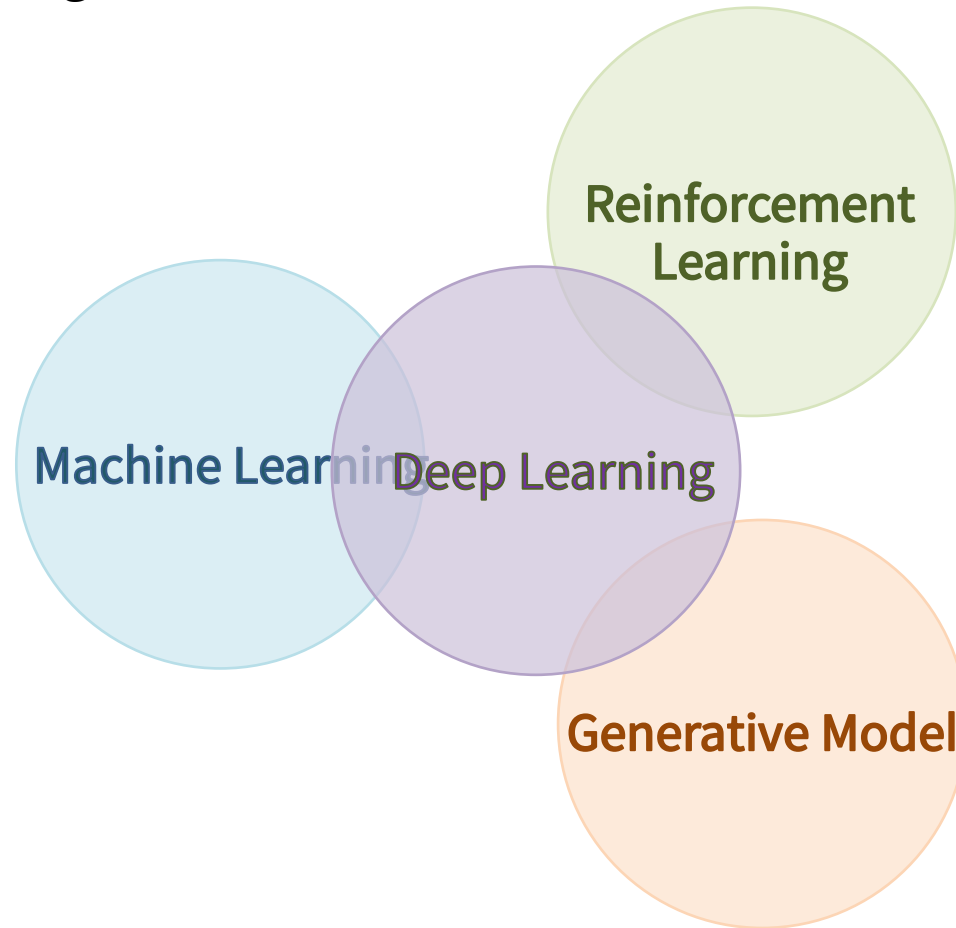
- 예술품을 생성해내는 GAN모형
 - 기존에 없었던 미술품을 생성하는 GAN모형인 Creative Adversarial Networks (CAN)



CAN의 결과물

| 인공지능 (Artificial Intelligence) 이란

인공지능 (Artificial Intelligence)



| 머신러닝의 개념과 정의

머신러닝(Machine Learning)의 정의

- 최근에 부르는 머신 러닝은 행과 열이 존재하는 행렬 (정형 데이터)을 가지고 무언가 예측 또는 분류 하고 싶을 때 사용한다라고 볼 수 있음
- 크게 보았을 때에 머신 러닝이 인공지능 안에 포함되는 개념이라고 볼 수 있지만, 비전문가나 대중들에게 이야기 할 때에는 용어를 혼용해서 사용하기도 함.

정형 데이터(행렬) - (Tabular data)

X						Y
id	레벨	경험치	게임시간	파티수	아이템정보	게임이탈여부
...

X: 게임 유저들의 게임 내 활동 정보, Y: 게임 이탈 여부 / 어뷰징 여부 -> 이상 탐지 모델

| 머신러닝의 개념과 정의

머신러닝(Machine Learning)의 정의

- “무엇(X)으로 무엇(Y)을 예측하고 싶다”

모델 (머신 러닝 알고리즘)

$$Y = f(X)$$

출력 변수
(종속변수, 반응변수)

입력 변수
(독립변수, feature)

- 주어진 데이터를 통해서 입력변수와 출력변수간의 관계를 만드는 함수 f 를 만드는 것
- 주어진 데이터 속에서 데이터의 특징을 찾아내는 함수 f 를 만드는 것

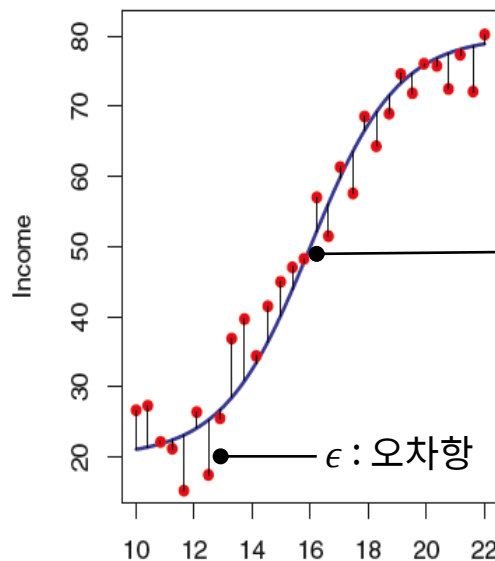
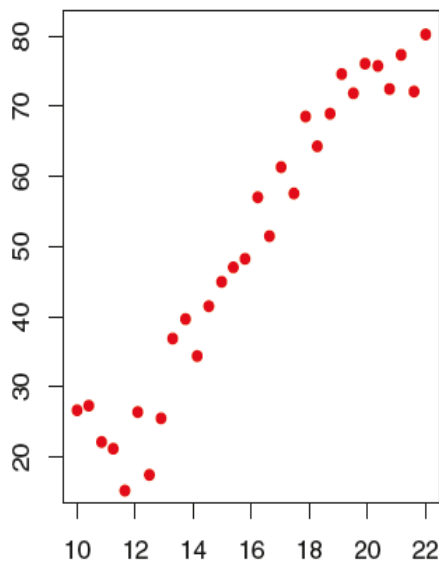
| 머신러닝의 개념과 정의

머신러닝(Machine Learning)의 정의

f 에 대하여

- p 개의 입력 변수 X_1, X_2, \dots, X_p 가 있고, 출력 변수 Y 가 있을 때, $X = (X_1, X_2, \dots, X_p)$ 라 하면 입력 변수와 출력 변수의 관계를 나타내는 식은 다음과 같음

$$Y = f(X) + \epsilon$$



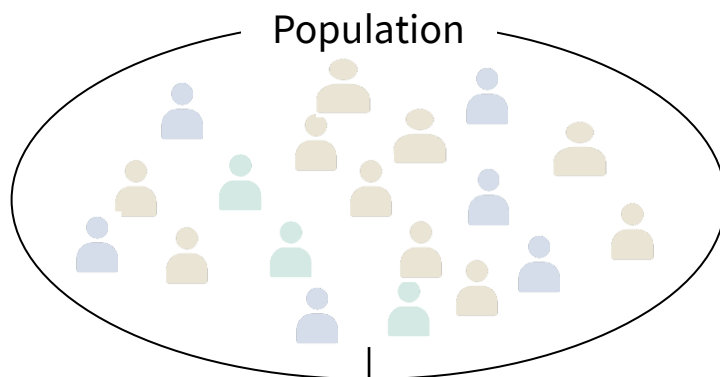
f : 정해졌지만 알 수 없는 함수

ϵ : 오차항

| 머신러닝의 개념과 정의

머신러닝(Machine Learning)의 정의

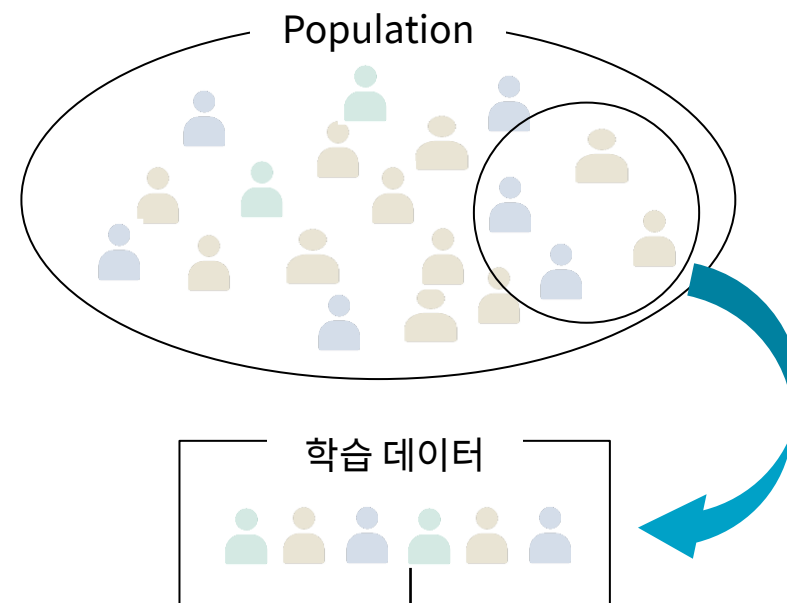
Population을 알 때



$$Y = f(X)$$

'신'만이 알고 있음!

Population을 모를 때



$$\hat{Y} = \hat{f}(X)$$

using machine learning
(statistics)

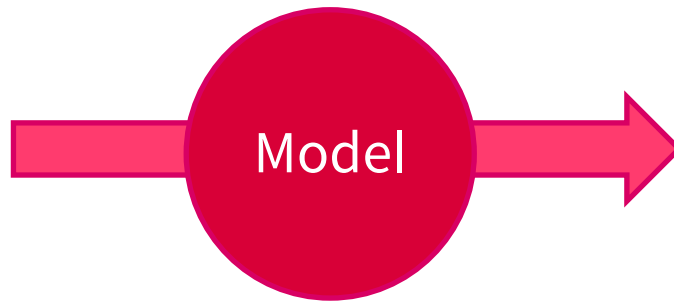
| 머신러닝의 개념과 정의

모델 학습

- 머신 러닝과 딥러닝의 공통 단어인 Learning, 즉 학습을 위해서는 가이드 혹은 정답이 필요함
- 머신 러닝 모델의 학습 목표는 데이터 (Input)에 대한 모델의 결과 (Output)가 정답 (Label)에 가깝게 나오게끔 학습시키는 것이라고 볼 수 있음.
- 강아지 사진 (Input)을 모델에 입력하면 강아지 (Output)로 분류하는 모델을 학습시키거나, 과거의 주가 데이터 (Input)을 입력하면 미래의 주가 데이터 (Output)를 예측하는 모델을 만드는 것



Input



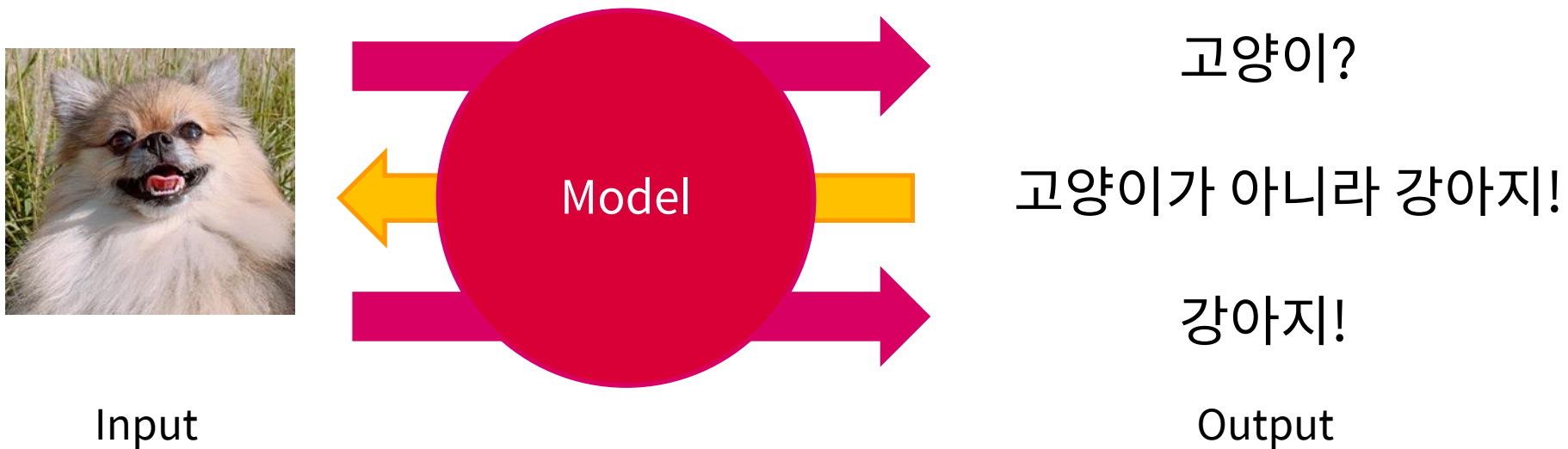
“강아지”

Output

| 머신러닝의 개념과 정의

모델 학습 과정

1. 데이터를 모델에 넣고, 결과를 낸다.
2. 결과를 정답과 비교해서 다른 만큼 모델을 변경한다.
3. 특정 조건이 만족할 때까지 1,2번을 반복한다.



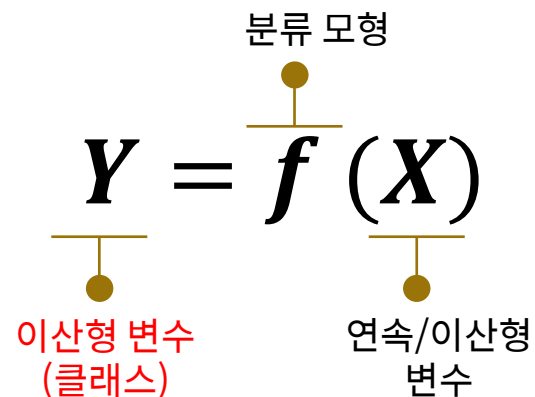
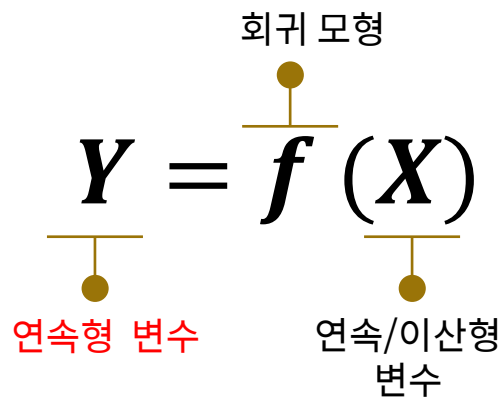
| 머신러닝의 구분과 종류

머신러닝의 구분

- 지도 학습(supervised learning)

$Y = f(X)$ 에 대하여 입력 변수 (X)와 출력 변수 (Y) 의 관계에 대하여 모델링하는것
(Y 에 대하여 예측 또는 분류하는 문제)

- 회귀 (regression): 입력 변수 X 에 대해서 연속형 출력 변수 Y 를 예측
- 분류 (classification): 입력 변수 X 에 대해서 이산형 출력 변수 Y (class)를 예측



| 머신러닝의 구분과 종류

머신러닝의 구분

- 비지도 학습(Unsupervised learning)

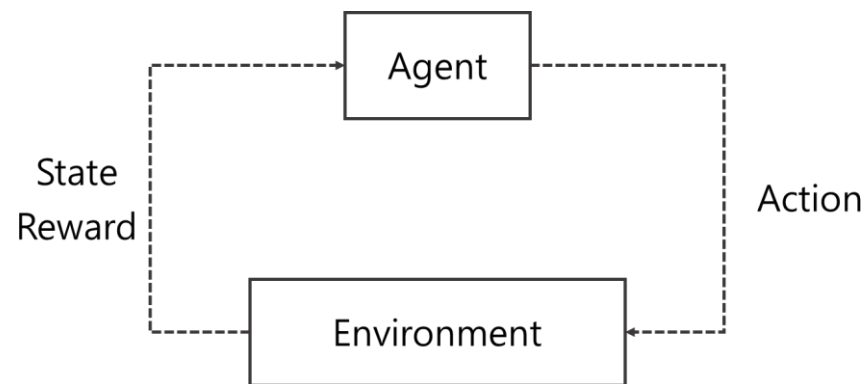
- 출력 변수(Y)가 존재하지 않고, 입력 변수(X)간의 관계에 대해 모델링 하는 것
- 데이터는 제공하지만 명확한 정답은 제공하지 않는 학습 방법
- 군집화, 추천시스템, 차원축소법 등

| 머신러닝의 구분과 종류

머신러닝의 구분

- 강화학습 (Reinforcement Learning)

- 강화학습은 수많은 시뮬레이션을 통해 컴퓨터가 현재 상태에서 어떠한 행동을 취해야 먼 미래의 보상을 최대로 할 것인가에 대해 학습하는 알고리즘
- 강화학습을 하기 위해서는 현재 상태 (State), 행동 (Action), 보상 (Reward), 다음 상태 (Next State)가 있어야 학습을 진행 할 수 있음.
- 기존의 머신 러닝처럼 데이터의 행이 각각 독립적으로 구성 되어있는 것이 아니라, 일련의 Episode가 있어야 함.

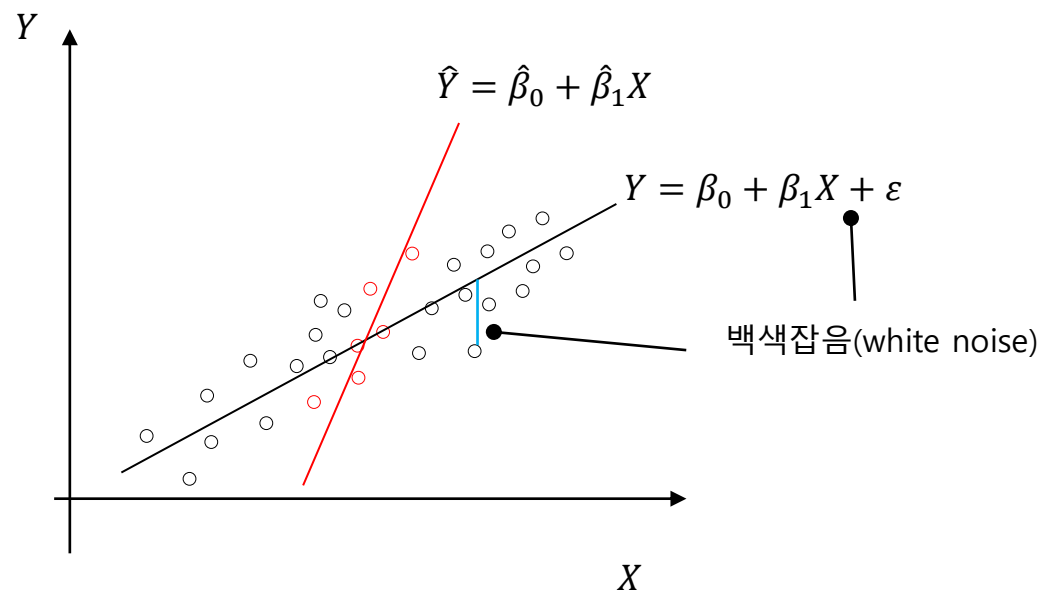


| 머신러닝의 구분과 종류

머신러닝의 종류

- 선형 회귀분석(Linear Regression)

- 독립변수와 종속변수가 선형적인 관계가 있다라는 가정하에 분석
- 직선을 통해 종속변수를 예측하기 때문에 독립변수의 중요도와 영향력을 파악하기 쉬움



| 머신러닝의 구분과 종류

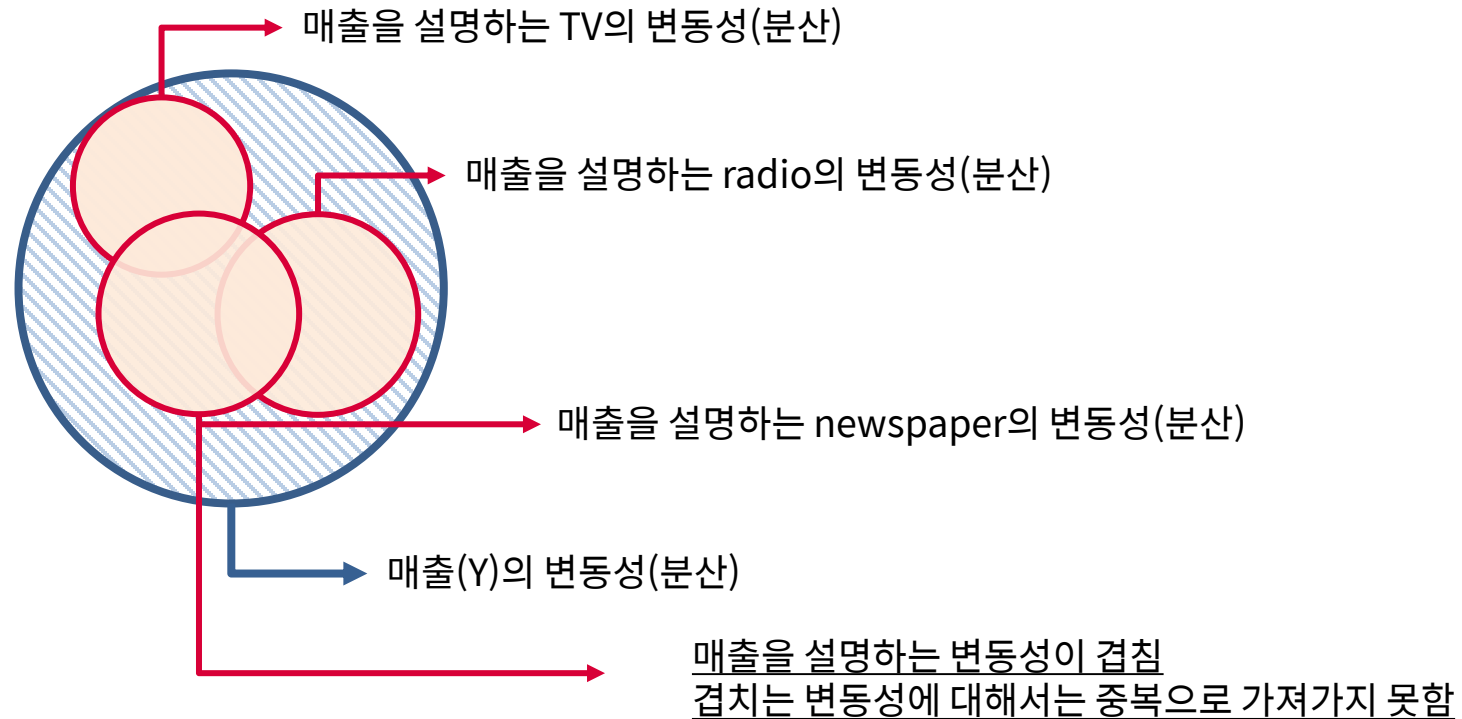
머신러닝의 종류

- 다중 선형 회귀 모델 (Multi Linear Regression Model)
 - 변수가 여러 개인 선형 회귀 모델
 - 선형 회귀 모델은 기본적으로 직선의 방정식과 같이 모델을 만들 수 있기 때문에, 가지고 있는 독립변수가 종속변수에 대하여 어떠한 형태로 영향을 끼치는 지 알 수 있음.
 - 즉, 보험사기, 대출연체, 게임이탈예측 문제를 생각 해 보았을 때 예측하는 것 자체로 물론 중요하지만, 어떠한 사람이 보험 사기를 행하는 어떠한 사람이 대출을 연체하는지 어떠한 사람이 게임이탈을 하는지 중요할 수도 있음.
 - 선형 회귀 모델은 성능 면에서 매우 우수한 알고리즘은 아니지만, 변수의 설명력에 있어서 강력한 장점을 가짐
 - 변수가 증가하면 다중공선성 문제 발생
-

| 머신러닝의 구분과 종류

머신러닝의 종류

- 다중 선형 회귀 모델 (Multi Linear Regression Model)



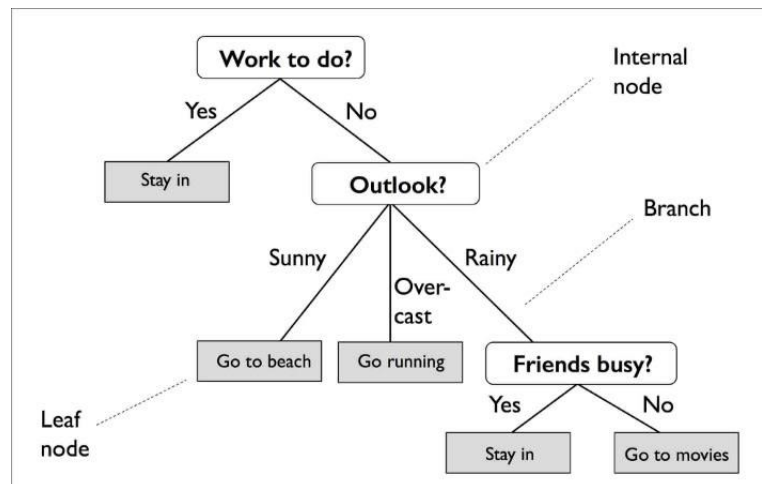
이러한 현상에 대해서 변수들간의 다중공선성(Multicollinearity)이 있다고 한다.
잘못된 변수해석, 예측 정확도 하락 등을 야기시킨다

| 머신러닝의 구분과 종류

머신러닝의 종류

- 의사 결정 나무 (Decision Tree)

- 선형 회귀 모델과 같이 매우 해석력이 높고 직관적인 모델
- 밝이 맑고 습도가 낮으면 테니스를 “한다”, 다만 습도가 높으면 “하지 않는다” 처럼 종속 변수가 독립 변수의 어떤 특정 조건에 의해 잘 나눌 수 있는가를 설명한 모델
- 설명력이 큰 장점이지만 Prediction의 성능은 부족한 부분이 있어서 단일 모델보다는 이를 발전시킨 Ensemble 모델들을 많이 사용

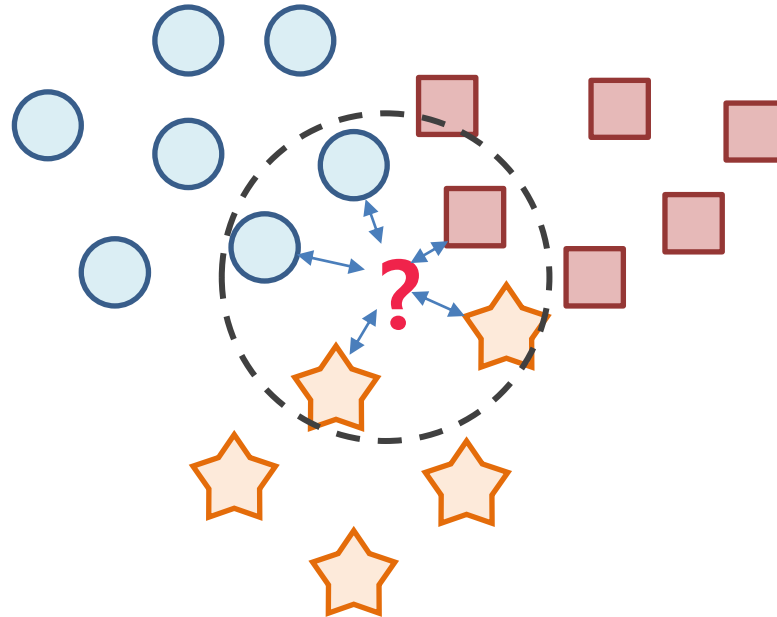


| 머신러닝의 구분과 종류

머신러닝의 종류

- k-Nearest Neighbor (kNN)

- 주어진 데이터의 가장 가까운 k개의 데이터를 통해 해당 데이터의 예측하는 직관적인 모델
- K는 사용자가 지정해줘야 하는 Hyperparameter로 사전에 지정해줘야 하며, 데이터간 거리 측정 지표나, k개 데이터의 정보를 종합하는 방법들을 선택하여 모델의 변화를 줄 수 있음

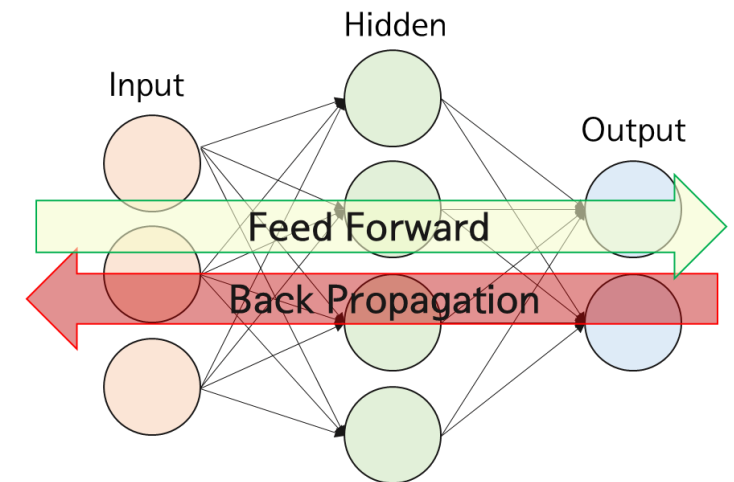
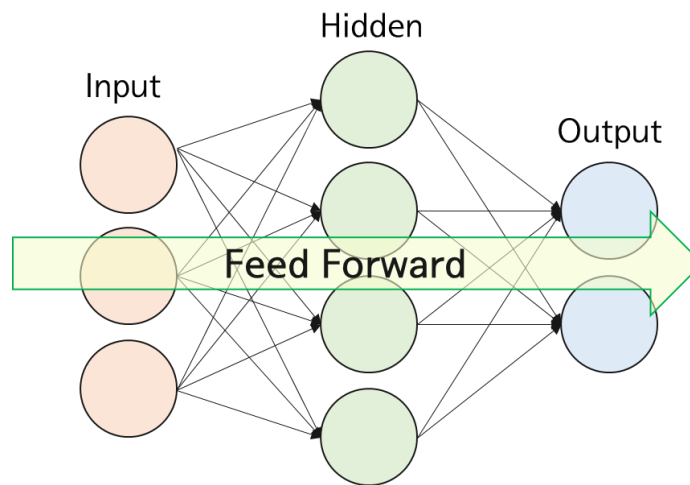
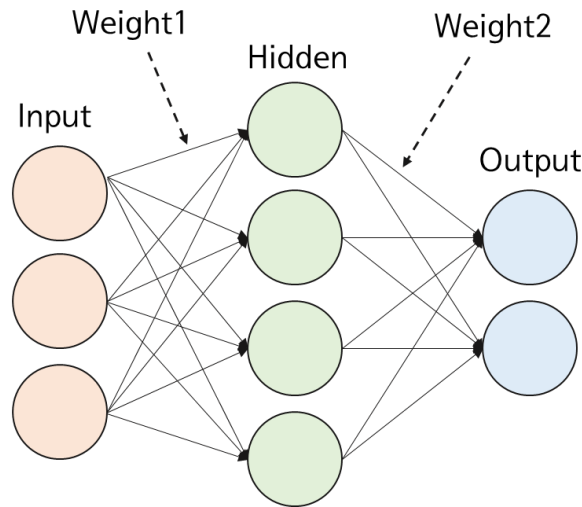


| 머신러닝의 구분과 종류

머신러닝의 종류

- 신경망 (Neural Network)

- 딥러닝의 가장 기초가 되는 모델로서 입력 (Input), 은닉 (Hidden), 출력 (Output)층으로 구성된 모형으로서 각 층을 연결하는 노드의 가중치를 Update하면서 학습
- 신경망의 output으로 부터 오차를 계산하고 오차를 줄여 나가면서 학습을 진행

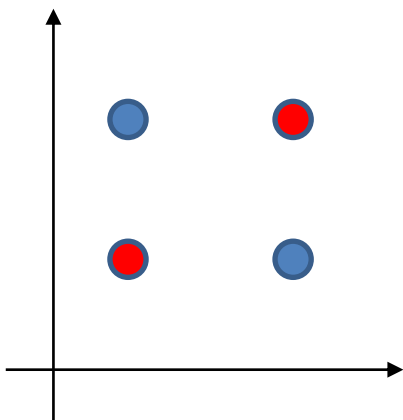


| 머신러닝의 구분과 종류

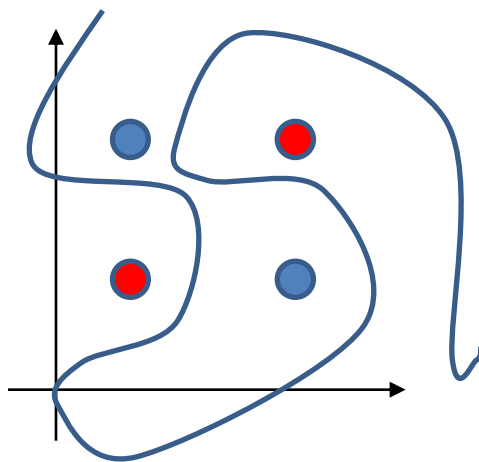
머신러닝의 종류

- 신경망 (Neural Network)

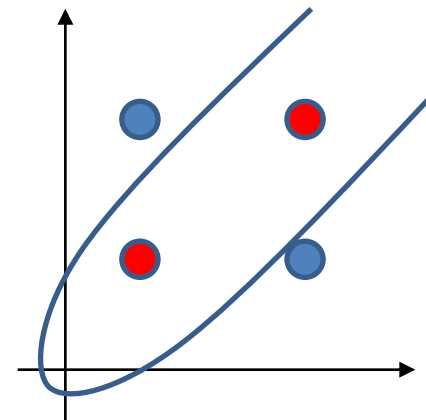
- 신경망은 학습 과정의 특성상 학습데이터에서 완벽한 모델을 만들 수 있음.
- 우리가 실제로 예측해야만 하는 데이터 (검증 또는 테스트 데이터)에 대해서는 잘 맞추지 못하는 현상이 발생 (과적합 현상)



풀고자 하는 문제



신경망으로 과적화 된 예시



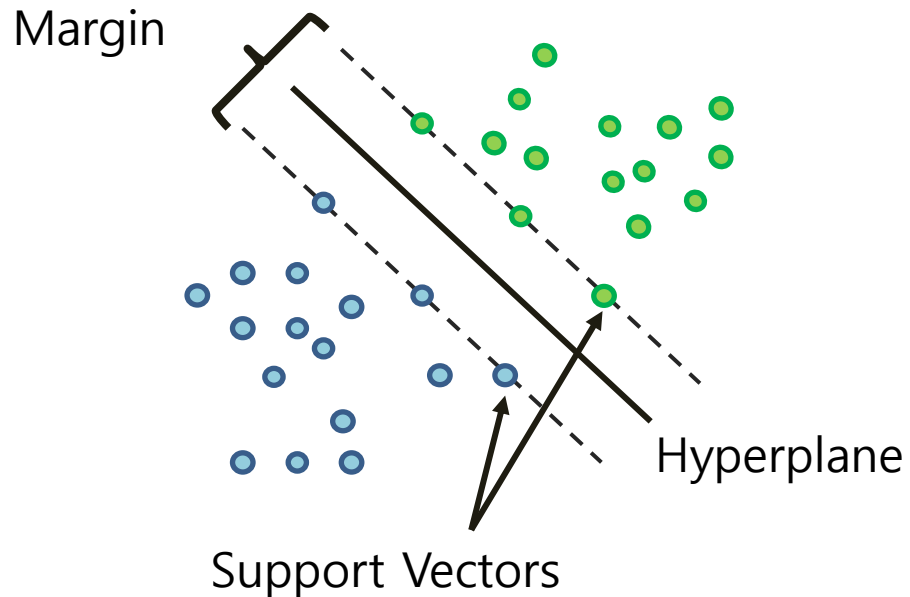
적절한 모형의 Decision Boundary

| 머신러닝의 구분과 종류

머신러닝의 종류

- Support Vector Machine (SVM)

- 직선을 그었을 때 class간의 거리가 각각 비슷하게끔, 즉 직선을 가운데로 잘 갖도록 하는 학습 방법을 가지고 있음
- 내가 가지고 있는 학습 데이터 내에서 일정 오류를 허용하고 직선을 그을 수도 있음
- 학습 과정 내에서 과적합을 어느정도 방지할 수 있는 알고리즘
- 데이터의 차원이 커질 수록 학습하는 시간이 매우 오래 걸린다는 단점

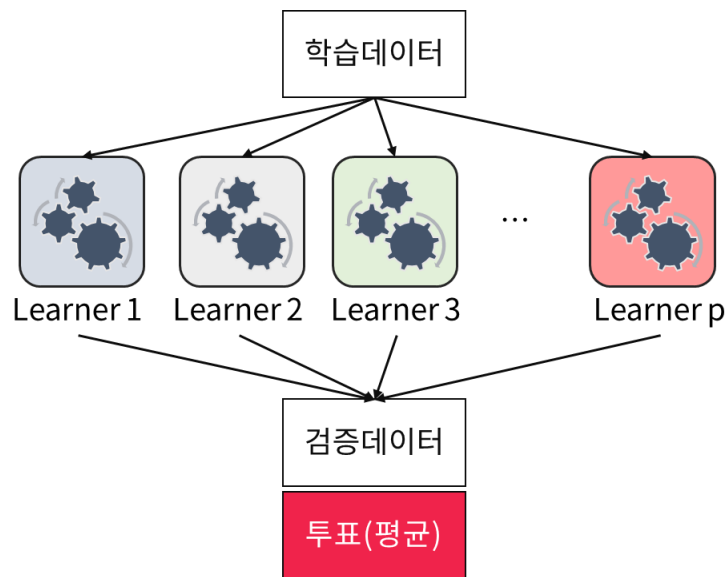


| 머신러닝의 구분과 종류

머신러닝의 종류

- Ensemble Learning

- 다양한 모델을 만들면 더 좋은 모델을 만들 수 있다는 컨셉. 다양한 모델을 만들어서 우리가 예측하고자 하는 모델에 대해 다양한 의견을 수렴해 투표를 통해 최종적인 예측 값을 만들고자 하는 것
- 데이터를 재구성해서 모델을 만드는 **Bagging**, 데이터 뿐만이 아니라 변수 또한 랜덤하게 추출해서 모델을 만드는 **RandomForest**, 잘 맞추지 못하는 데이터를 더 집중적으로 학습시키는 **Boosting** 등이 있음

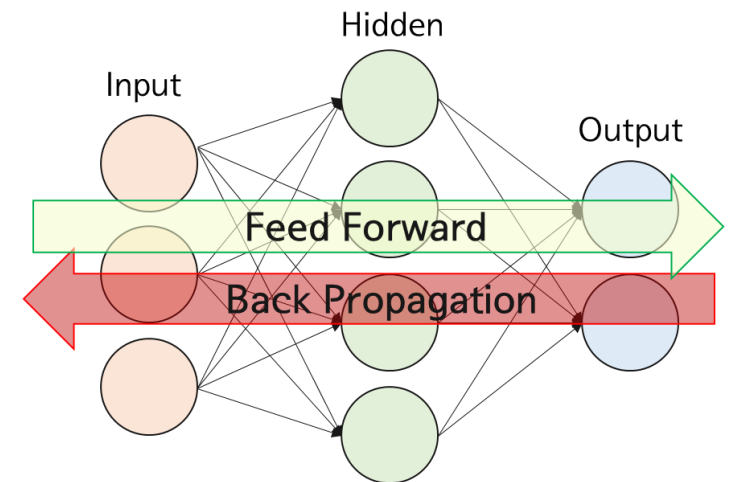
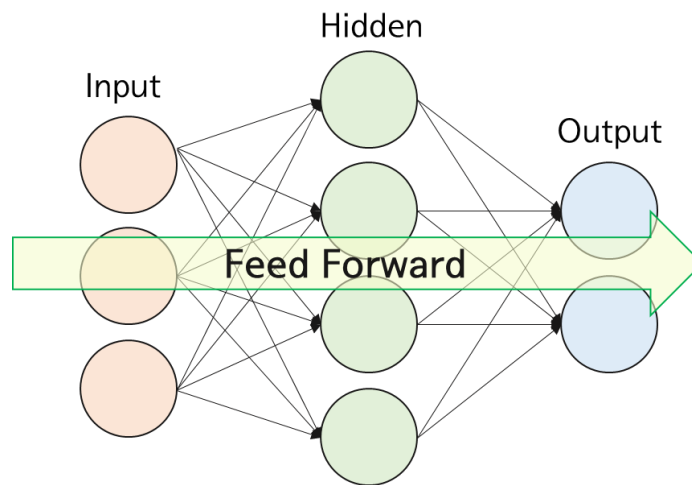
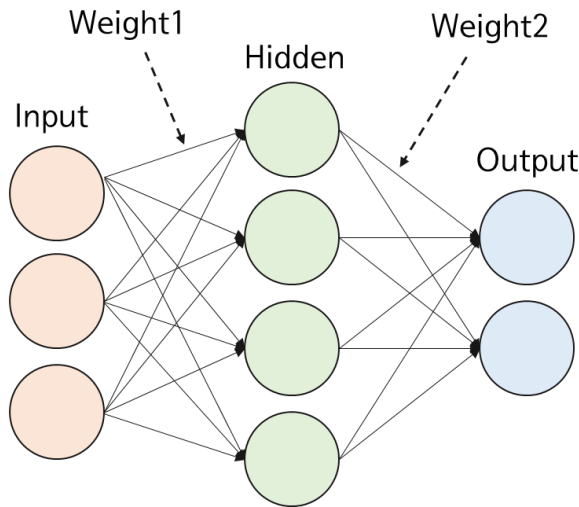


| 과적합 (Overfitting)

과적합

- 신경망 (Neural Network)

- 딥러닝의 가장 기초가 되는 모델로서 입력 (Input), 은닉 (Hidden), 출력 (Output)층으로 구성된 모형으로서 각 층을 연결하는 노드의 가중치를 Update하면서 학습
- 신경망의 output으로 부터 오차를 계산하고 오차를 줄여 나가면서 학습을 진행

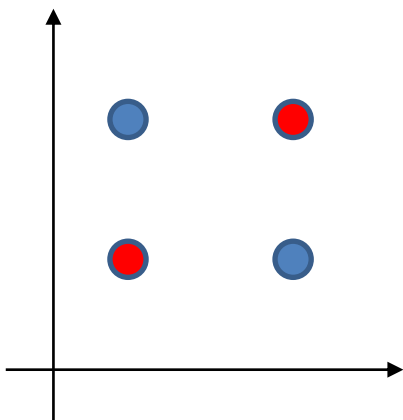


| 과적합 (Overfitting)

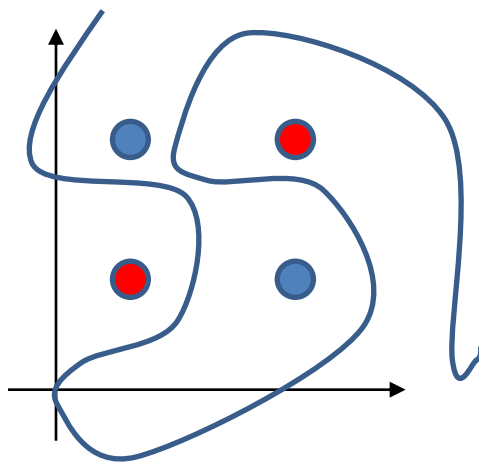
과적합

- 신경망 (Neural Network)

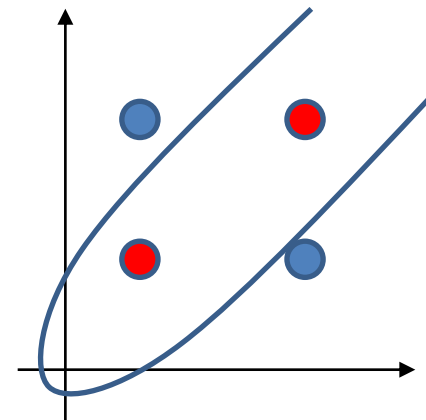
- 신경망은 학습 과정의 특성상 학습데이터에서 완벽한 모델을 만들 수 있음.
- 우리가 실제로 예측해야만 하는 데이터 (검증 또는 테스트 데이터)에 대해서는 잘 맞추지 못하는 현상이 발생 (과적합 현상)



풀고자 하는 문제



신경망으로 과적화 된 예시

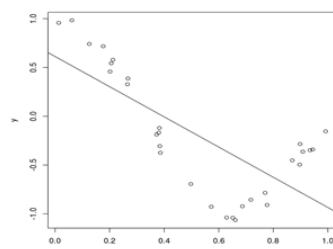


적절한 모형의 Decision Boundary

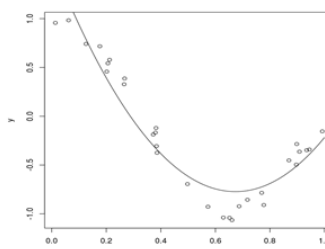
| 과적합 (Overfitting)

과적합

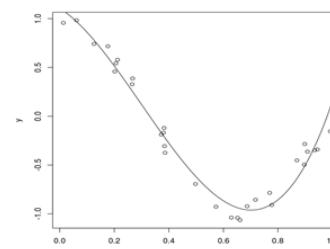
- 우리가 가지고 있는 샘플(학습) 데이터를 완벽하게 맞추는 모델을 만든다면, 우리가 보지 못한 데이터에 대해 잘 맞는 다라는 보장이 있을까?
- 우리가 가지고 있는 데이터를 더욱 더 완벽하게 맞추려고 노력한다면 우리가 보지 못한 데이터는 더 맞추지 못할 확률이 높음



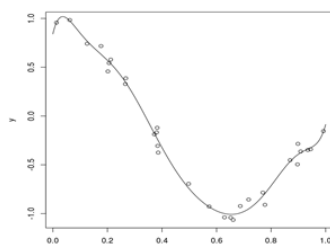
1차항 고려



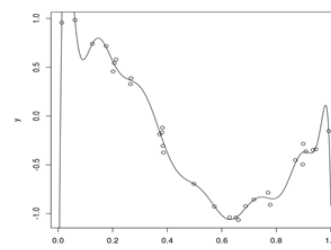
2차항 고려



3차항 고려



4차항 고려 (과적합 발생)



5차항 고려 (과적합 발생)

선형 회귀 모델의 과적합 예시

| 과적합 (Overfitting)

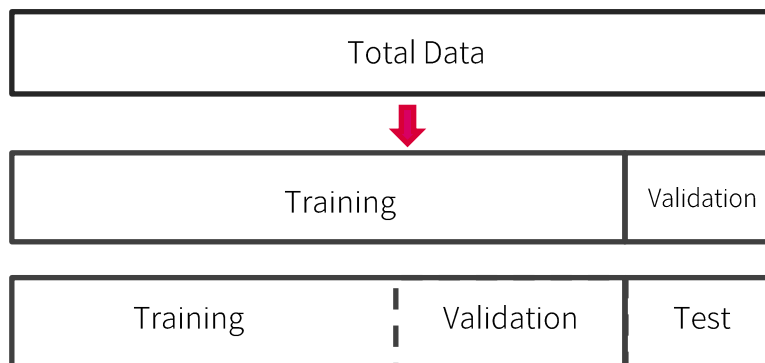
과적합

- 학습 할 샘플 데이터 수의 부족 : 학습 할 데이터의 수가 많으면 많을수록 모집단의 특성을 잘 반영할 확률이 높으며, 우리가 가지고 있는 데이터가 모집단의 특성을 잘 반영할 수록 과적합이 일어날 확률이 적어진다
- 풀고자 하는 문제에 대비 복잡한 모델을 적용 : 간단한 문제에 대해서는 간단한 모델 또는 적은 변수만을 사용하는 것이 과적합 될 가능성을 줄일 수 있습니다. 모델이 복잡하다 라는 표현은 많은 변수를 사용하는 것 또는 모델 자체가 복잡한 모델 (SVM, 딥러닝 등)이라는 것을 의미

| 과적합 (Overfitting)

과적합을 방지하는 방법

- 데이터의 수를 늘려 모집단의 특성을 잘 반영한 데이터를 확보
- 풀고자 하는 문제에 대하여 적절한 모형을 선택
- 적절한 실험설계
 - 데이터를 학습/검증/테스트 데이터로 랜덤하게 분할
 - 학습 데이터 (Training Data) : 머신 러닝 모델 함수 f 를 적합 시킬 데이터
 - 검증 데이터 (Validation Data) : 학습 데이터를 적합 시킨 함수 f 를 검증 시킬 데이터
 - 테스트 데이터 (Test Data) : 테스트 데이터는 최종적으로 우리가 만든 모델의 성능을 측정하는 데이터임. 검증 데이터와 테스트 데이터의 다른 점은, 테스트데이터는 우리가 전혀 보지 못한 데이터라고 가정 해야 한다라는 것



| 과적합 (Overfitting)

적절한 실험설계

- K-fold cross validation
 - 전체 데이터를 K개의 Fold로 랜덤하게 데이터를 분할
 - 첫번째 Fold를 제외한 나머지 Fold 데이터를 합쳐서 학습 데이터로 사용하고 첫번째 Fold를 검증 데이터로 사용
 - 두번째 Fold를 검증 데이터로 사용하고 두번째 Fold를 제외한 나머지 Fold 데이터를 학습 데이터로 사용
 - 이를 K번 반복하면 결국에, 모든 데이터를 학습 데이터로서 사용하면서 검증 데이터로서 사용 할 수 있음



...



Training

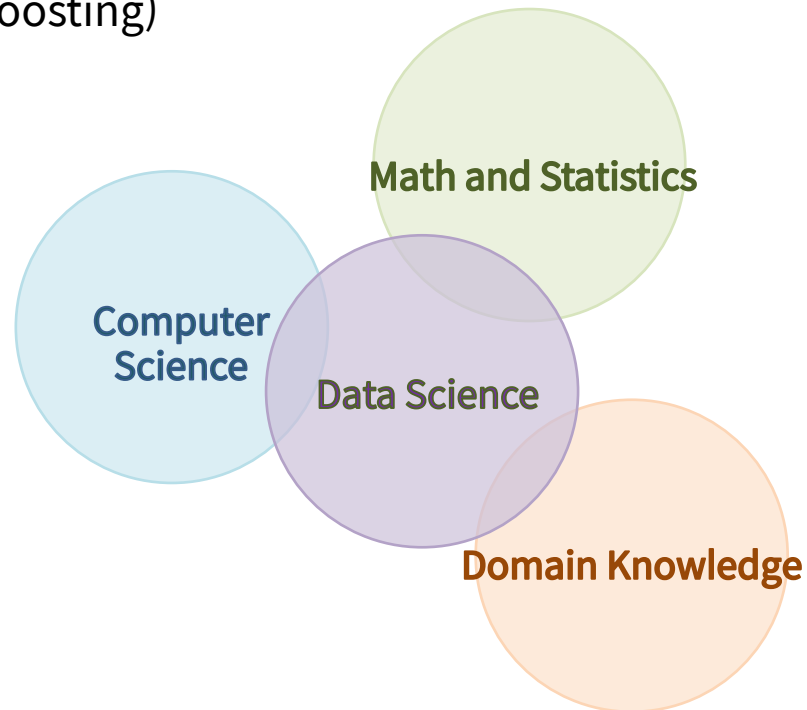


Validataion

DataScience

DataScience

- Data analysis, Machine learning, Data Mining
- Data를 학습시키고 정보를 추출(Feature Extraction) 또는 예측하는 분야(Machine Learning)
- 대부분의 Machine Learning알고리즘은 Feature Extraction을 동시 수행 (Lasso/Ridge, RandomForest, boosting)



| DataScience

고려해야할 것들

- 1. 어떠한 모델을 쓸 것인가
(Linear Regression, Decision Tree, K-nearest Neighbor, Support Vector Machine, Ensemble Learning, Deep Learning)
 - 2. Y의 영향을 끼치는 좋은 변수가 있는가
(데이터 전처리 및 파생 변수 생성)
 - 3. 최적의 변수의 조합을 고려했는가(최종적으로 사용하는 변수는 어떤 변수인가)
(Feature Selection)
 - 4. 실험 설계를 제대로 했는가
(과적합 문제)
 - 5. 모델을 통해서 얻을 수 있는 것은 무엇인가
(Insight)
-