



머신러닝의 개요

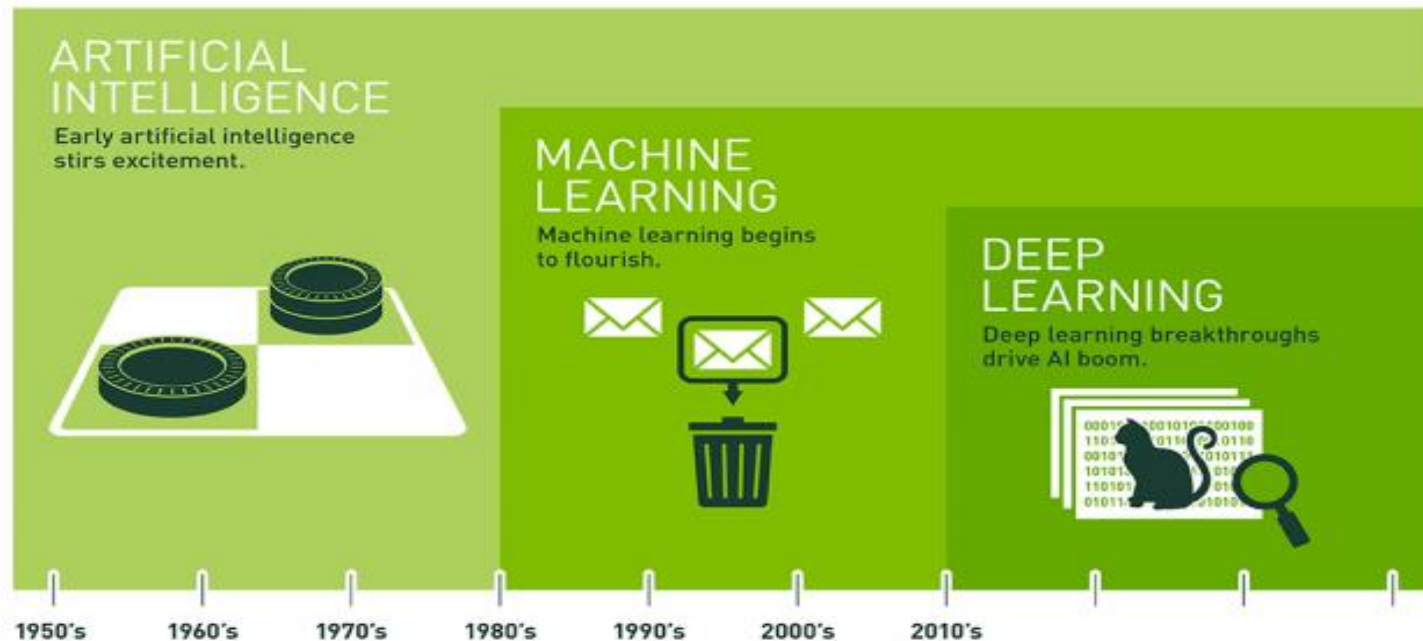
안 화 수

머신러닝

❖ 머신러닝의 개념

머신러닝(Machine Learning)은 인공지능의 한 분야로 기계 스스로 대량의 데이터로부터 지식이나 패턴을 찾아 학습하고 예측하는 알고리즘 기법을 통칭한다.

Deep Learning은 Machine Learning의 일종이고, Machine Learning은 인공지능의 일종이다.



머신러닝

❖ 머신러닝의 분류

학습 데이터에 레이블(label)이 있는 경우와 그렇지 않은 경우에 따라 지도 학습과 비지도 학습으로 구분하고, 강화 학습은 지도 학습 중 하나로 분류되거나 또는 독립적인 세 번째 머신러닝 모델로 분류하기도 한다.

Types	Tasks	Algorithms
지도 학습 (Supervised Learning)	분류 (Classification)	<ul style="list-style-type: none">▪ KNN : k Nearest Neighbor▪ SVM : Support Vector Machine▪ Decision Tree (의사결정 나무)▪ Logistic Regression
	예측 (Prediction)	<ul style="list-style-type: none">▪ Linear Regression (선형 회귀)
비지도 학습 (Unsupervised Learning)	군집 (Clustering)	<ul style="list-style-type: none">▪ K-Means Clustering▪ DBSCAN Clustering▪ Hierarchical Clustering (계층형 군집)
강화 학습 (Reinforcement Learning)		<ul style="list-style-type: none">▪ MDP : Markov Decision Process

머신러닝의 용어

- ❖ **회귀(Regression)**
과거의 데이터를 기반으로 미래의 데이터를 예측하는 것이다.
판매 예측, 주가 변동 등을 예측하는 경우를 의미한다.
- ❖ **분류(Classification)**
특정 데이터에 레이블을 붙여서 분류할 수 있다. 예를 들면, 스팸메일 분류, 필기 인식, 증권 사기 등에 사용하는 경우를 의미한다.
- ❖ **군집화(Clustering)**
값의 유사성을 기반으로 데이터를 여러 그룹으로 나눌 수 있다. 예를 들면 사용자의 취향을 그룹으로 묶어 사용자 취향에 맞는 광고를 제공하는 경우를 의미한다.
- ❖ **추천(Recommendation)**
특정 데이터를 기반으로 다른 데이터를 추천하는 것이다. 예를 들면 사용자가 인터넷 서점에서 구매한 책들을 기반으로 다른 책을 추천하는 경우를 의미한다.
- ❖ **차원 축소(Dimensionality Reduction)**
데이터의 특성을 유지하면서 데이터의 양을 줄이는 것이다. 데이터의 특성을 유지한 상태로 고차원의 데이터를 저차원의 데이터로 변환하는 것이다. 데이터를 시각화하거나 구조를 추출해서 용량을 줄여 계산을 빠르게 하거나 메모리를 절약할 때 사용한다.

초과학습(Overfitting)

❖ 초과학습, 초과적합(Overfitting)이란?

“ 초과학습(Overfitting) ”이란 훈련 전용 데이터는 잘 학습되어 있지만, 학습되지 않은 새로운 데이터에 대해서는 제대로 예측하지 못하는 상태를 의미 한다.

❖ 초과학습의 원인

1. 학습 데이터가 너무 적은 경우
2. 모델에 비해서 문제가 너무 복잡한 경우

❖ 초과학습의 해결 방법

1. 학습 데이터가 적은 경우에 근본적인 해결 방법은 데이터 수를 늘리는 것이다.
2. 모델에 비해서 문제가 너무 복잡한 경우에는 다른 모델을 선택해서 적용해야 한다.
3. 교차검증 (Cross Validation)
4. Dropout

파이썬 기반의 머신러닝

- ❖ 파이썬 기반의 머신러닝 생태계를 구성하는 패키지
 - 머신러닝 패키지 : 사이키런(Scikit-Learn)
 - 행렬/선형대수/통계 패키지 : Numpy
 - 데이터 핸들링 패키지 : Pandas
 - 시각화 패키지 : Matplotlib, Seaborn