

# 시계열 데이터분석

Timeseries Data Analytics

LG전자 CDO부문

정의석 연구원

02

# 머신러닝 & 딥러닝

# 1

## 인공지능, 머신러닝, 딥러닝의 관계

인공지능  
Artificial Intelligence

머신러닝  
Machine Learning

딥러닝  
Deep Learning

인공지능 ⊃ 머신러닝 ⊃ 딥러닝

인공지능 (AI) :

사람의 지적 능력을 컴퓨터를  
통해 구현하는 기술

머신러닝 (ML) :

데이터로부터 의사결정을 위한  
패턴을 기계가 스스로 학습

딥러닝 (DL) :

의사결정을 위해 기계가  
데이터로부터 숨겨진 특징을  
추출하고 학습

## ❖ 머신러닝 (Machine Learning)

- 데이터를 통해 패턴을 찾고, 이를 바탕으로 예측이나 의사결정을 할 수 있음.

## ❖ 딥러닝 (Deep Learning)

- 머신러닝의 한 분야로, 인공 신경망(Artificial Neural Network)을 기반으로 함.
- 다층 퍼셉트론과 같은 복잡한 구조를 가진 신경망을 사용하여 데이터의 숨겨진 특징을 찾을 수 있음.

## ❖ 공통점

- 데이터 기반: 둘 다 데이터를 기반으로 학습하고 예측이나 분류 작업을 수행함.
- 패턴 인식: 둘 다 데이터에서 패턴이나 특징을 추출하여 사용함.

## ❖ 차이점

- 구조 복잡성: 딥러닝은 일반적으로 더 복잡한 구조(다층 신경망)를 가짐.
- 데이터 양: 딥러닝은 학습에 대량의 데이터가 필요, 머신러닝은 상대적으로 적은 데이터로도 가능함.
- 학습 데이터: 머신러닝은 주로 정형 데이터를, 딥러닝은 주로 비정형 데이터를 다룸.
- 학습 시간: 딥러닝은 일반적으로 머신러닝보다 더 많은 계산 리소스와 시간이 필요함.
- 해석 가능성: 머신러닝 모델(특히, 선형 모델, 의사결정 트리 등)은 딥러닝 모델보다 해석하기 쉬움.

### ❖ 지도학습 (Supervised Learning)

- 입력 데이터와 레이블을 함께 사용하여 모델을 학습시킴.
- $f(x) = y$  형태의 함수를 찾는 것이 목표임.
  - ✓ 예시: 선형 회귀 (Linear Regression), 의사결정나무 (Decision Tree)

### ❖ 비지도학습 (Unsupervised Learning)

- 레이블이 없는 입력 데이터만을 사용하여 모델을 학습시킴.
- 데이터의 숨겨진 구조나 패턴을 찾는 것이 목표임.
  - ✓ 예시: K-평균 군집화 (K-means Clustering), 생성 모델 (Generative Models)

### ❖ 반지도학습 (Semi-supervised Learning)

- 지도학습과 비지도학습을 혼합하여 사용함.
- 일부만 레이블이 있는 데이터를 사용하여 모델을 학습시킴.
  - ✓ 예시: 레이블 전파 (Label Propagation), 셀프 트레이닝 (Self-training)

### ❖ 강화학습 (Reinforcement Learning)

- 에이전트가 환경과 상호작용하며 보상을 최대화하는 행동을 학습함.
- $Q(s, a)$  함수를 사용하여 상태  $s$ 에서 행동  $a$ 를 취했을 때의 기대 보상을 계산함.
  - ✓ 예시: Q-러닝 (Q-Learning), 딥 Q 네트워크 (Deep Q-Network, DQN)

# 4 주요 성능 측정 지표 : 회귀와 분류

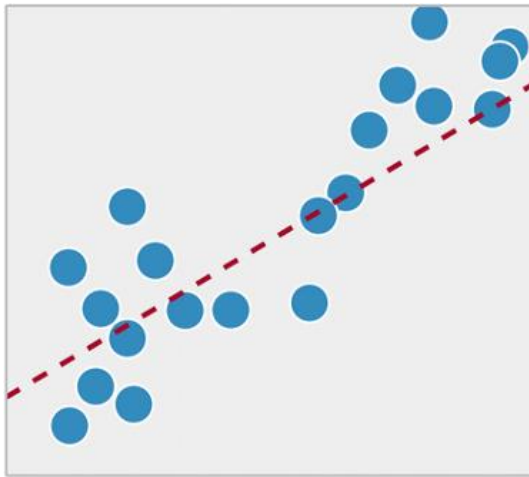
❖ 예측하고자 하는 값에 따라 회귀와 분류로 나눌 수 있음

❖ 회귀 (Regression)란?

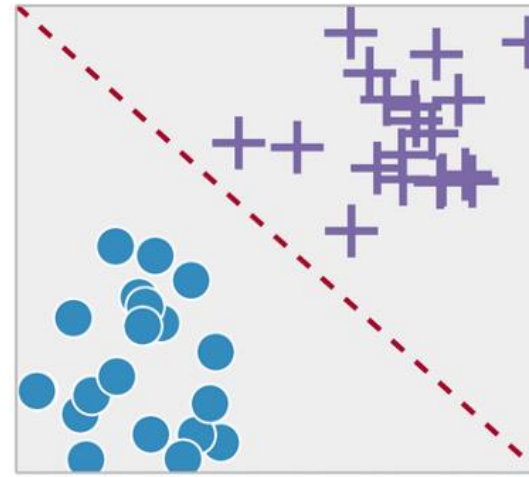
- 목적: 연속적인 수치 값을 예측하는 것이 목적임.
- 예시: 주택 가격 예측, 온도 예측, 매출액 예측 등

❖ 분류 (Classification)란?

- 목적: 레이블이 있는 카테고리로 데이터를 분류하는 것이 목적임.
- 예시: 스팸 메일 필터링, 이미지 분류, 질병 진단 등



Regression Task



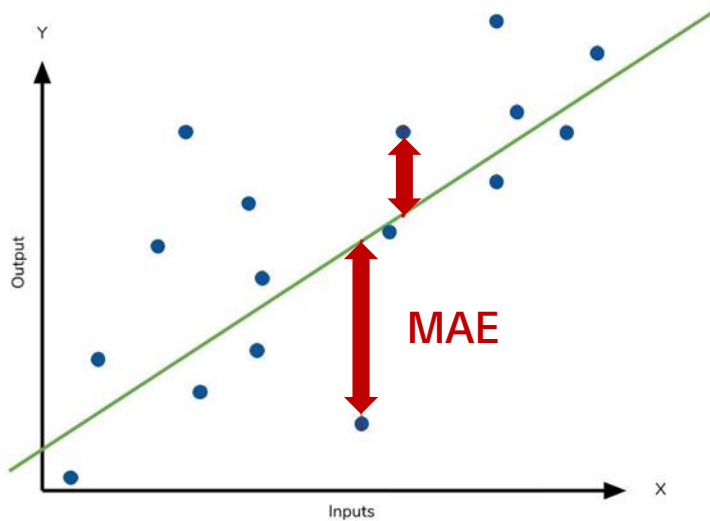
Classification Task

# 4

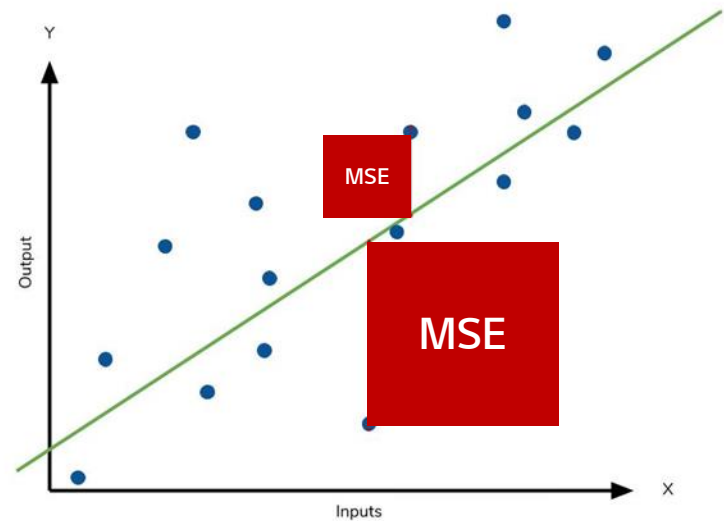
## 주요 성능 측정 지표 : 회귀와 분류

### ❖ 회귀 (Regression) 성능 측정 지표

- Mean Squared Error (MSE) : 예측 값과 실제 값 차이의 제곱 평균값  $\frac{1}{N} \sum_i^N (Y'_i - Y_i)^2$
- Root Mean Squared Error (RMSE) : MSE의 제곱근 값  $\sqrt{\frac{1}{N} \sum_i^N (Y'_i - Y_i)^2}$
- Mean Absolute Error (MAE) : 예측 값과 실제 값 차이의 절대값 평균 값  $\frac{1}{N} \sum_i^N |Y'_i - Y_i|$



Mean Absolute Error



Mean Square Error

# 4

## 주요 성능 측정 방법론 : 회귀와 분류

### ❖ 분류 (Classification) 성능 측정 지표

#### • 정확도 (Accuracy)

- ✓ 모델이 올바르게 분류한 샘플의 비율

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Samples}$$

#### • 정밀도 (Precision)

- ✓ 모델이 양성이라고 예측한 것 중 실제 양성 비율

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

#### • 재현율 (Recall)

- ✓ 실제 양성 중 모델이 양성이라고 예측한 비율

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

#### • F1 스코어 (F1-Score)

- ✓ Precision과 Recall의 조화평균 값

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Precision



실제 \ 예측	True	False
True	True Positive	False Positive
False	False Negative	True Negative

Recall



실제 \ 예측	True	False
True	True Positive	False Positive
False	False Negative	True Negative



# 4

## 주요 성능 측정 방법론 : 회귀와 분류

### ❖ 분류 (Classification) 성능 측정 지표

#### • 정확도 (Accuracy)

- ✓ 모델이 올바르게 분류한 샘플의 비율

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Samples}$$

#### • 정밀도 (Precision)

- ✓ 모델이 양성이라고 예측한 것 중 실제 양성 비율

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

#### • 재현율 (Recall)

- ✓ 실제 양성 중 모델이 양성이라고 예측한 비율

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

#### • F1 스코어 (F1-Score)

- ✓ Precision과 Recall의 조화평균 값

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Precision



실제 \ 예측	True	False
True	True Positive	False Positive
False	False Negative	True Negative

Recall



실제 \ 예측	True	False
True	True Positive	False Positive
False	False Negative	True Negative

## ❖ 데이터 세트란?

- 데이터 세트란 데이터들의 집합체 즉, 자료들의 모음을 일컬음.

## ❖ Train(학습), Validation(검증), Test(테스트) Set (\*기계 = 고등학생)

- 훈련 세트(train set)는 기계(학생)를 학습시키기 위한 교과서
- 검증 세트(validation set)은 잘 학습하고 있나 중간중간 확인할 수 있는 모의 고사
- 테스트 세트(test set)은 학습된 모델의 최종 성능을 확인하기 위한 수능 시험

## ❖ Train(7)과 Test(3)만 사용하는 경우

- 데이터셋의 크기가 작거나, 모델의 복잡도가 낮아 과적합(Overfitting)의 위험이 적을 때 사용함.
- 파라미터 튜닝이 필요하지 않거나, 튜닝이 간단할 때 사용함.

## ❖ Train(6), Validation(2), Test(2) 사용하는 경우

- 데이터셋의 크기가 크거나, 모델의 복잡도가 높아 과적합의 위험이 있을 때 사용함.
- 파라미터 튜닝이 복잡하거나, 여러 번의 튜닝이 필요할 때 사용함.

**감사합니다**