



# 데이터 분석 기반 AI 시스템 개발자 양성 과정

## 06.인공지능 구현 with python

조창제

강의자료

2025.08.

# 목차

## CONTENTS



I

데이터의 분포적 특성을 파악하는

### 탐색적 자료분석

II

결측치 대체 및 데이터 변환

### 데이터 전처리

III

적합한 모델 선정 및 모델 검증

### 모델 생성 및 검증





## 01. 개요

### 1. 워크플로우

- ① 데이터 수집
  - 다양하고 균형있는 데이터 수집이 중요
  - 일관된 형식을 유지하는 것이 중요
- ② 데이터 탐색
  - 이상치 탐지 및 데이터 분포 특성 파악
- ③ 데이터 전처리
  - 결측치에 대한 적절한 처리
- ④ 모델 선택 및 학습
  - 모델 최적화 및 다양한 모델 비교
- ⑤ 모델 평가
  - 적합한 평가 지표 선택



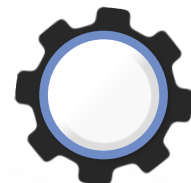
데이터 수집



데이터 탐색



데이터 전처리



모델 선택 및 학습



모델 평가



### 1. 정의

#### 1) 탐색적자료분석(Exploratory data analysis, EDA)

① 주어진 데이터셋에 대해 시각화하거나 통계적 요약을 통해 패턴, 특성, 구조를 파악하는 초기 분석 과정

- 중심 경향과 산포정도 파악 필요
- 이상치 및 누락된 값을 식별 및 해결 필요

② 데이터 구조 파악

- 기초통계량
  - 데이터의 분포와 패턴 파악 용이
  - 데이터 변동성 및 데이터의 대칭성 파악

③ 이상치 탐지 및 처리

- 분위수(Quantile)
- 사분위범위(Interquartile Range, IQR)



대시보드 시각화 예시



### 1. 정의

#### 1) 탐색적자료분석(Exploratory data analysis, EDA)

##### ① 시각화

- 변수 간 관계
  - 상관행렬(Correlation matrix)

##### ② 분포 시각화

- 히스토그램(Histogram)
  - 연속형 데이터를 구간별로 나누어 데이터의 빈도나 수를 시각적으로 나타낸 그래프
- 상자그림(Box plot)
  - 사분위수를 기반으로 데이터를 요약하여 보여주는 그래프
  - 이상치 탐지에 용이
- 산점도(Scatter plot)
  - 두 개 이상의 변수 사이의 관계를 시각적으로 표현하는 그래프
  - 변수 간 상관성이나 분포를 직관적으로 파악 가능



대시보드 시각화 예시



### 1. 정의

#### 1) 변수변환

- ① 데이터의 분포나 범위를 변경하는 과정을 의미
  - 정규화: 0~1사이 범위로 변환하는 과정
  - 표준화: 평균이 0 표준편차가 1인 분포로 변환하는 과정
  - 로그변환, boxcox변환: 비정규 분포를 가질 때, 정규분포에 가깝게 변환

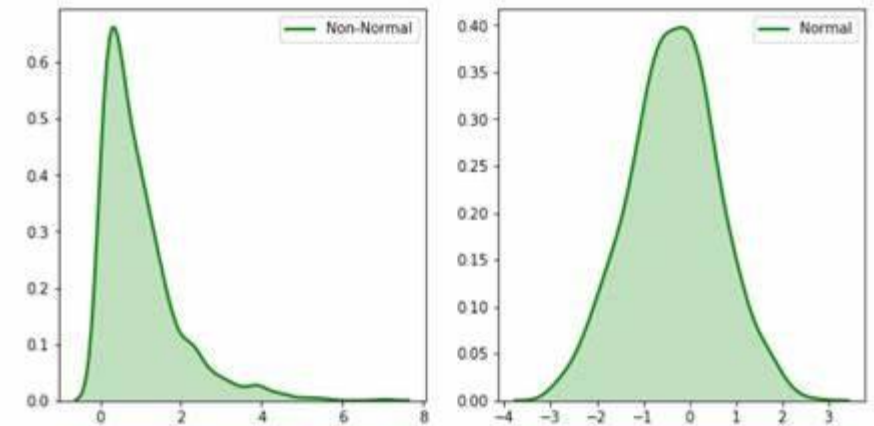
#### 2) 결측값 처리

- ① 결측보간: 관측되지 않거나 누락된 값을 채워넣는 기법
  - 결측보간보다 해당 데이터를 제거하는 것이 더 좋을 수도 있음

#### 3) 파생변수 생성

- ① 기존 변수를 활용하여 다양한 방식 변수를 생성
  - 예시 - 키, 몸무게를 활용하여 BMI 변수 생성
  - 예시 - 날짜 및 시간을 sin, cos 등의 주기를 고려한 변수 생성

Lambda value used for Transformation: 0.30656155175590766



Boxcox 변환 예시



### 1. 정의

#### 1) 알고리즘 선택

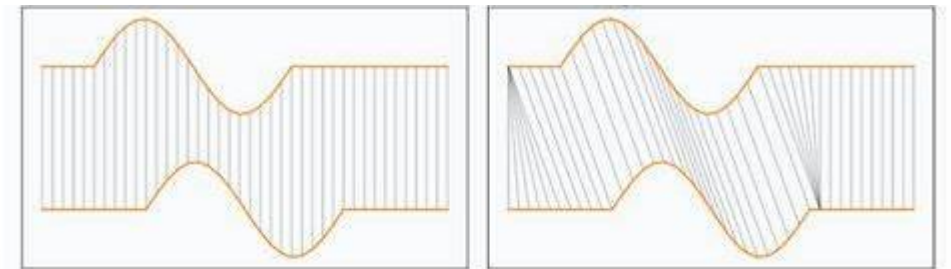
##### ① 데이터 불균형

- SMOTE 방식 활용
  - 작은 레이블의 데이터 비중을 늘려 비율이 유사하게 만들어 내는 기법
- Class Weight 적용
  - 학습 시 loss 계산에서 레이블에 대한 가중치를 적용

#### 2) 목적함수

##### ① 성능평가지표

- 적합한 성능평가 지표 설정 필요
- 시계열 모델의 검증에서 MAE나 MSE는 부적합 할 수도 있음
- 이미지 모델 성능평가에서 MAE 또한 부적합 할 수 있음



Dynamic time warping 예시



### 1. 정의

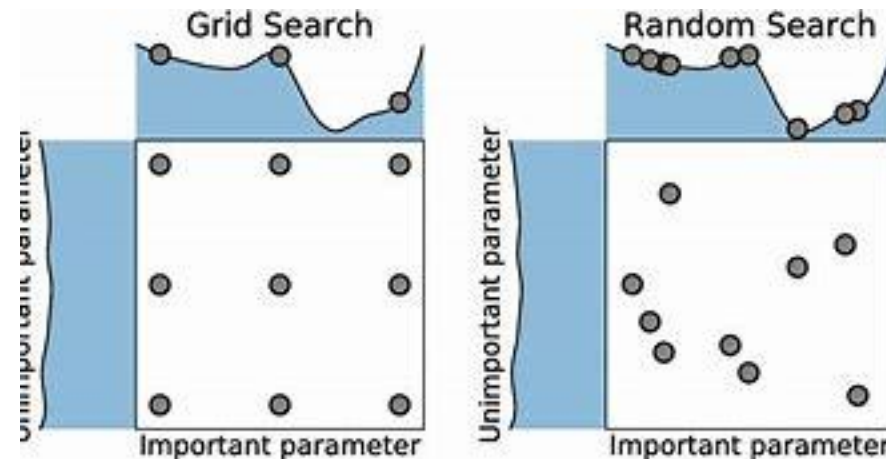
#### 1) 하이퍼 파라미터 최적화

- ① 효율적인 최적화 방식 적용 필요

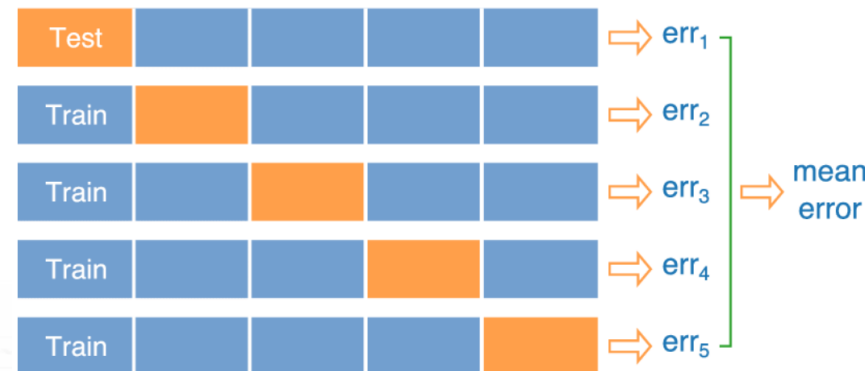
#### 2) 검증 방식

##### ① 교차검증

- K fold cross validation
- Hold out cross validation
- leave one out cross validation



Hyper parameter tuning 종류



K fold cross validation





# Thank You

Email: [qkdrk777777@naver.com](mailto:qkdrk777777@naver.com)