| 교육 제목 | 데이터 기반 인공지능 시스템 엔지니어 양성 과정\_판다스 |
| --- | --- |
| 교육 일시 | 2021년 10월 5일 |
| 교육 장소 | YGL C-6 학과장 & 자택(디스코드 이용한 온라인) |
| **교육 내용** | |
| 오전 | Review  파이썬 머신러닝 판다스 데이터 분석 5장 \_ 데이터 사전 처리   1. 누라데이터 처리 : 확인 -> 제거(dropna()) 또는 치환(fillna(method=’ffill’/’bfill’)) 2. 중복 데이터 처리 : 확인 -> 제거(drop\_duplicates()) 3. 데이터 표준화 : 단위 환산, 자료형 변환(astype()) 4. 범주형(카테고리) 데이터 처리 : 군간 분할 -> 더미변수(원핫인코딩, get\_dummies()) 5. 정규화 : 숫자 데이터의 상대적인 크기 차이를 제거 6. 시계열 데이터 : 주식 환율 등 금융 데이터를 다루기 위해 개발된 기능    * 다른 자료형을 시계열 객체로 변환 - 문자열을 Timestamp로 변환 : to\_datetime() 함수로 datetime64 자료형으로 변환 - Timestamp를 Period로 변환 : to\_period() 함수와 freq옵션을 이용 변환    * 시계열 데이터 만들기 - Timestamp 배열 : date\_range() 함수를 이용  - Period 배열 : period\_range() 함수를 이용    * 시계열 데이터 활용 - 날짜 데이터 분리 :  - 날짜 인덱스 활용 : 연-월-일 중 내가 필요로 하는 레벨을 선택적으로 인덱싱 가능 |
| 오후 | 파이썬 머신러닝 판다스 데이터 분석 6장 \_ 데이터프레임의 다양한 응용   1. 함수 매핑 - apply() 메소드 : 인자로 전달하는 매핑 함수에 시리즈의 모든 원소를 하나씩 입력하고 함수의 리턴값을 돌려받는다. (같은 크기의 시리즈 객체로 반환) - applymap() 메소드 : 데이터프레임의 각 원소를 매핑 함수에 하나씩 넣어서 리턴값으로 돌려받음. (동일한 형태의 데이터프레임으로 반환)    * 시리즈 객체에 함수 매핑 - 데이터프레임의 각 열에 함수 매핑 : df.apply(매핑함수, axis=0) : 모든 열을 하나씩 분리하여 매핑 함수의 인자로 각 열(시리즈)에 전달됨. - 데이터프레임의 각 행에 함수 매핑 : df.apply(매핑함수, axis=1) : 각 행을 매핑 함수의 인자로 전달    * 데이터프레임 객체에 함수 매핑 - pipe() 메소드 활용 : 매핑 함수에 따라 반환되는 객체의 종류가 달라짐 2. 열 재구성    * 열 순서 변경 : df[재구성한 열 이름의 리스트] - sorted() 함수: 알파벳 순으로 정렬 - reversed() 함수: 역순으로 정렬    * 열 분리 : 문자열 데이터화 후 Series 객체.str.get(인덱스) 3. 필터링    * 불린 인덱싱 : df[불린 시리즈]    * isin() 메소드 활용 : df의 열 객체.isin(추출값의 리스트) 4. 데이터프레임 합치기    * 데이터프레임 연결 : pd.concat(데이터프레임의 리스트)    * 데이터프레임 병합 : pd.merge( 옵션들…)    * 데이터프레임 결합 : df.join(df2, 옵션들…) 5. 그룹 연산    * 그룹 객체 만들기 - 1개 열을 기준으로 그룹화 : df.groupby(기준이 되는 열) - 여러 열을 기준으로 그룹화 : df.groupby(기준이 되는 열의 리스트)    * 그룹 연산 메소드 (적용-결합 단계) - 데이터 집계 :  !!! 내장함수 이용 (표준편차 구하기) group객체.std()  !!! 사용자 정의 함수 이용 : agg() 메소드 : group객체.agg(매핑함수) - 그룹 연산 데이터 변환 : group객체.transform(매핑함수) - 그룹 객체 필터링 : group객체.filter(조건식함수) - 그룹 객체에 함수 매핑 : group 객체. apply(매핑 함수) 6. 멀티 인덱스 : 판다스는 멀티 인덱스를 지원함 7. 피벗 : 판다스는 엑셀의 피벗테이블과 비슷한 기능을 처리   파이썬 머신러닝 판다스 데이터 분석 7장 \_ 머신러닝 데이터 분석  데이터 분석이란? 예측을 위한 수식을 만드는 것이다.  데이터 분석의 진행 순서   1. 자료 수집 2. 누락데이터 처리 및 데이터 전처리(데이터 가공, 문자더미,NaN처리, 분류화, 조인/분리/결합…) 3. 실험데이터와 test 데이터 분리 4. 모델 학습 5. 모델 결과 타겟값 계산 6. 테스트 데이터를 모델값에 적용한 결과와 실제 테스트 타겟값 비교 7. 점수 확인 8. 그래프화 : 예측된 값과 실제값 비교 9. 모델 검증 |