| 교육 제목 | 데이터 기반 인공지능 시스템 엔지니어 양성 과정\_판다스 |
| --- | --- |
| 교육 일시 | 2021년 10월 1일 |
| 교육 장소 | YGL C-6 학과장 & 자택(디스코드 이용한 온라인) |
| **교육 내용** | |
| 오전 | 파이썬 머신러닝 판다스 데이터 분석 5장 \_ 데이터 사전 처리   1. 누라데이터 처리    * 누락 데이터 확인 - df.info() 메소드 : 각 열에 속하는 데이터 중에서 유효한 값의 개수 확인 - value\_counts() 메소드 : dropna=False 옵션으로 누락데이터 개수 확인 - isnull() or notnull() 메소드 : 누락데이터를 찾는 직접적인 방법    * 누락데이터 제거 - dropna() 메소드 : thresh=000 옵션으로 NaN 000 개 이상의 열 삭제  subset을 특정열로 한정하고 axis=0 how=any 옵션으로 NaN 포함 행 삭제    * 누락데이터 치환 - fillna() 메소드 : 특정값으로 치환, 또는 앞뒤값(method=’ffill’ or ‘bfill’ 옵션) inplace=True 옵션 꼭! 사용 - NaN 값이 아닌 것을 치환 : df.replace(‘?’, np.nan, inplace=True) 2. 중복 데이터 처리    * 중복 데이터 확인 - duplicated() 메소드 : 중복이면 True 반환    * 중복 데이터 제거 - drop\_duplicates() 메소드 : inplace=True 옵션 추가,subset=[,,,] 로 특정열지정 3. 데이터 표준화    * 단위 환산 : 측정 단위를 동일하게 조정    * 자료형 변환 : 숫자가 문자열로 저장된 경우 숫자형으로 변환 또는 그 반대  !! 유한 개의 고유값이 반복 → astype(‘category’) 메소드로 범주형으로 변경 4. 범주형(카테고리) 데이터 처리    * 구간 분할  - cut() 함수 옵션으로 분할 x=df[‘배열할 열']  ! bins= : 경계값 리스트, labels= : 구간 이름 리스트, include\_lowest=True (낮은 경계값 포함)    * 더미 변수 : 컴퓨터가 인식 가능한 입력값으로 변환  one hot vector 로 변환 시키기 때문에 one-hot-encoding 이라고도 함  - get\_dummies() 함수를 사용하여 더미 변수로 변환 5. 정규화 : 숫자 데이터의 상대적인 크기 차이를 제거 6. 시계열 데이터 : 주식 환율 등 금융 데이터를 다루기 위해 개발된 기능 |
| 오후 | 파이썬 머신러닝 판다스 데이터 분석 6장 \_ 데이터프레임의 다양한 응용   1. 함수 매핑    * 개별 원소에 함수 매핑 - apply() 메소드 : 시리즈의 모든 원소를 하나씩 입력하고 리턴값을 돌려 받음    * 데이터프레임 원소에 함수 매핑 - applymap() 메소드 : 데이터프레임의 각 원소를 하나씩 입력하고 리턴값을 돌려받음 2. 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 3. 데이터프레임 객체에 함수 매핑    * pipe() 메소드 활용 |