|  |
| --- |
| **SSAC**  **파이썬을 활용한 비즈니스 빅데이터 분석가 양성 과정**  **(TEMPLATE) 자가학습 워크시트** |

|  |  |
| --- | --- |
| **이름** | 백승재 |
| **날짜** | 2021-09-15 |
| **온라인 콘텐츠(인강)**  **학습 범위** | * 직장인을 위한 Python : 0203~0205 * Python : Numpy 04\_05 ~ 05\_02복습 |
| **복습 노트 작성 (자유 양식)** | 02-02.list, 02-03.tuple, 02-04.set, 02-05.dic   * 1. List (중복가능, 순서, 수정가능),순서가 있어 Sequence Data   - [], append, remove(제거 앞), pos/neg indexing, len   * 2. Tuple (중복가능, 순서, 수정 불가능, 읽기 전용), 순서가 있어 Sequence Data   - (), len   * 3. SET(중복 x, 순서x , 집합)   - {},set(), set.add()   * 4. Dic (중복 x, key(수정불가능), value로 이루어진 사전형 집합)   - {},dict(),dict(key) = value  04-05. Array function, 04-06 Array Aggregation functions   * Argmax, argmin 설정 축에 최소 argument의 위치(index)를 추출 * Cumsum : index순으로 누적합, Cumprod: index순으로 누적곱 * Np. Sort 기본 행순으로 정렬한다. * Argsort: 정렬된 값의 원래 위치를 추출 * Np.empty와 np.zeros는 초기화하는 함수로 메모리를 할당하고 0으로 초기화하는 것은 zero, 0으로 초기화하지 않는 것은 empty로 수행속도는 empty가 조금 더 빠르다.   04-07. Performance Test   * Numpy가 파이썬 보다 기본 1000배 정도 수행 속도가 빠름 * Numpy는 Single Type이라 type matching 안함 * Numpy는 메모리에 Universal function을 로드해놓았기 대문에 function dispatching 을 안함   05\_01. Pandas란   * 정형데이터를 다루는 특화된 툴 (Structured Dat, PanelData, Tabular Data) * 다양한 라이브러리(Scikit-learn, scipy, statsmodel, tensorflow, pytorch)에 의존성 * Excel(피벗기능)과 sql 기능 사용 가능하면서 고성능 데이터 처리에 적합   05\_02. Pandas의 기본 자료 구조(Series, DataFrame)   * DataFrame 2차원 테이블, 테이블의 한줄(행/열)을 series라고 한다. * Series 생성 : pd.Series([]), indexs, values tables * 날짜 범위 생성 : Pd.date\_range(년도, 구간(periods)), type(datatype64) * Dataframe 생성 : index변수 만들기,np.random,randn value생성, column 명(리스트 부여) |
| **퀴즈 제출** | Q1. 다음 A\_set(type set)에 값을 추가하는 것으로 적절한 것은?  A\_set = {3,4,5}  1. A\_set.append(3)  2. A\_set[0] = 3  3. A\_set.add(4)  4.A\_set.add(6)  정답 : 4  Set type은 중복이 안되고 (3번은 값 추가 안됨), set은 순서가 없으며, 추가할 때는 add method를 사용한다. Append list 추가 method이며, 2번은 dict type에서 원소를 추가하는 방식이다.  Q2 . Mat Array를 이루는 모든 원소를 오름차순으로 정렬하는 것으로 적절한 것은?  Mat =([[0.24605913, 0.30482292, 0.33102963],  [0.39890301, 0.43628836, 0.49886564],  [0.56931493, 0.70178652, 0.70854956],  [0.7505077 , 0.83385551, 0.85224302],  [0.85790954, 0.91533382, 0.98654989]])   1. np.argsort(mat3, axis=1) 2. np.sort(mat(1,15), axis=0) 3. np.argsort(mat3, axis=0)   4. np.sort(mat, axies= 1)  정답 : 2  mat shape로 sort를 하면 row기준으로 정렬하기 때문에 형변환을 해준후 sort를 한다. sort함수만 default값(axis=0), 즉 row(행)순이고 나머지 함수에서는 axis=0은 column 기준이다. Argmax는 축기준 argument sorting한 것의 원래 위치를 보여준다.  Q3. 다음 아래 코드에 대한 설명으로 옳은 것은?  A= np.full((3,2),5), B =np.ones((2,3))  1. A.size는 6이고 B,size는 3이다.  2. A-3은 연산이 불가능한 코딩이다.  3. A.shape은 (3,3)이다.  4. A.dot(B)는 shape이 (3,3)이고 모든 요소는 10인 array가 된다.  정답 : 4 A.size & B.size 는 6, A-3은 Elemental wise연산이 가능한 코딩, A.shape(3,2)이다. |
| **.** |  |
| **화면 캡쳐**  **(교육 화면 + 교육일시)** |  |