|  |  |
| --- | --- |
| 교육 제목 | **데이터 다루기(train,test) , 데이터 전처리, 결정계수R2** |
| 교육 일시 | 2021.10.15 |
| 교육 장소 | 비대면 |
| **교육 내용** | |
| 오전 | **데이터 전 처리**  데이터 준비 -> 사이킷런 으로 train(훈련)세트 와 test(테스트) 세트 나누기  데이터 불러오기 : \_\_ =[] 형태 , type, np.shape로 타입과 개수 확인가능  **-knn을 이용한 분류**    **-zip 파이썬 제공함수 사용**    **-타겟 만들기**  **0과1로 값을 나타나게 함** bream(도미)의35개 값은 1로 ,smelt 14개의 값은0으로      **---샘플링---**  **-np.column\_stack ()** 넘파이 제공함수 사용  전달받은 리스트를 일렬로 세운 다음 차례대로 나란히 연결  -타겟 만들기  원소가 하나인 리스트 을 여러 번 곱해서    -train , test 나누기 |
|  | **---샘플링 편향---**  **-KNeighborsClassifier** = 주어진 샘플에서 가장 가까운 이웃을 찾아주는 메서드 제공    -데이터를 랜덤하게 바꾸기      **-np.ones() , np.zeros()** 각각 원하는 개수의 1과 0을 채운 배열을 만들어 줌    **- train\_test\_split()** 전달되는 리스트나 배열을 비율에 맞게 훈련 ,테스트 세트로 나눔    -predict([[ ]]) 안을2 차원으로 하고 안의 값을 데이터와 비교하여 결과 도출  ->0 값 나옴 |
|  | **-np.mean()** 평균값 계산  **-np.std()** 함수의 표준편차 계산    - 일반적 연산은 불가하지만 컴퓨터가 매트릭스 생성 후 연산 |
|  | **회귀문제Knr**  농어(perch) 문제로 예시  데이터 불러오기 -> 훈련,테스트세트로 split ->  - reshape 메서드 배열의 크기를 지정    -**KNeighborsRegressor :** 회귀 알고리즘을 구현한 클래스  객체 생성후 fit() 메서드로 회귀모델 훈련    **-결정계수 R^2**    **- mean\_absolute\_error :** 회귀모델의 평균 절대값 오차를 계산    **- mean\_squared\_error:** 평균 제곱 오차 계산    선형회귀  최근접 이웃의 한계 |