|  |  |
| --- | --- |
| 교육 제목 | 머신러닝 |
| 교육 일시 | 2021년 10 15일 |
| 교육 장소 | 영우글로벌러닝 |
| **교육 내용** | |
| 오전 | \* 마켓과 머신러닝 (생선 분류 문제, 도미데이터로 선점도 그리기, scatter() 함수를이용하여 선점도를 그림 zip() 함수를와 리스트 내포 구문을사용해서 리스트를 2차원리스트로 만들기 , 사이킷런 패키지에서 K-최근점 이웃 알고리즘을 구현한 클래스인  KNeighborsClassifier를 이용 객체만들기, fit() 메서드를 이용 머신러닝에서 훈련시킴  fit() 는 사이킷런 모델을 훈련할때 사용함, predict()는 사이킷런 모델을 훈련하고 예측할때 사용하는 메서드, score()는 훈련된 사이킷런 모델의 성능을 측정합니다.  \* 샘플링 편향이란 만약 도미와 빙어가있는데 훈련세트에 도미만 넣고 테스트세트에 빙어만 들어있으면 한쪽으로 치우치기때문에 정확도가 0으로 뜰수밖에 없다. 그러므로 골고루 섞어줘야된다.  \*이런 작업을 하기위해 numpy 모듈을 사용하면 쉽게 해결된다. 파이썬 리스트로는  2차원 리스트만 표현가능하나 numpy로는 고차원의 배열을 손쉽게 만들고 조작할수있다.  \*배열에서 랜덤하게 샘플을 선택하여 훈련세트와 테스트 세트를 만들고 배열을 섞은후 무작위로 샘플을 고를시에는 target값을 설정하여 구분해주어야 나중에 임의의 값을 입력해서 넣었을때 그값이 어느쪽에 속하는지 알수있다.  \* random패키지 아래있는 shuffle() 함수로 배열을 무작위로 섞을수 있다 |
| 오후 | \* 데이터 전처리: 넘파이의 column\_stack() 함수는 전달받은 리스트를 일렬로 나란히  세운 다음 차례대로 나란히 연결합니다. 타깃 데이터를 만들때 원소가 하나인 리스트  [1], [0]을 여러번 곱해서 타깃 데이터를 만들었는데 넘파이는 np.ones()와 np.zeros()  함수를 사용하면 각각 원하는 개수의 1과0을 채운 배열을 만들어 줍니다.  \* 사이킷런으로 훈련세트와 테스트 세트를 나눕니다. 대표적인 도구가 train\_test\_split()  함수입니다. 리스트나 배열을 비율에 맞게 훈련 세트와 테스트 세트로 나누어 줍니다.  \* 준비한 데이터로 K-최근점 이웃을 훈련해 보면 새로운 데이터를 넣었을때 단위가 x축 과 y축 차이가 많이 나면 정확한 데이터를 얻지못한다.  \*그러하므로 제대로 사용하려면 특성값을 일정한 기준으로 마추어주어야한다. 널리사용되는게 표준점수이다. 계산하는방법은 간단하다. 평균에서 값을 뺴고 표준편차를 나누어 주면 됩니다. 넘파이에선 np.mean() <평균> , np.std() <표준편차> 를사용함녀 됩니다. 그후에 산점도를 그리면 정확하게 k-최근점 이웃의 값을 산출 할수있습니다 |