|  |  |
| --- | --- |
| 교육 제목 | 머신러닝 회귀 |
| 교육 일시 | 10월 19일(화) |
| 교육 장소 | 재택 |
| **교육 내용** | |
| 전체 내용 | - 로지스틱 회귀  분류의 문제에 회귀를 적용함. 선형 회귀로는 현실을 예측하고 문제를 해결할 수 없다. 이러한 것을 해결하기 위해 고안된 것이 로지스틱 회귀이다. 로지스틱 회귀의 가장 큰 특징은 시그모이드 함수를 사용한다. 시그모이드 함수는 입력값(x)이 음수 방향으로 향할수록 출력값(y)이 0에 가까워지고, 입력값(x)이 양수 방향으로 향할수록 출력값(y)이 1에 가까워진다. 출력값은 무조건 0~1 사이에 존재한다. 기존 데이터셋의 y(출력값)를 Sigmoid function을 겨쳐 0과 1 사이의 정수로 나타내고 Pred\_y(예측값)을 구하고, 예측값이 어느 임계치를 넘는다면 True(1), 아니라면 False(0)을 반환하도록 한다.  - 지니지수 : Gini index  impurity(불순도) : 얼마나 불확실한가? 를 보여주는 지표이기 때문에 0에 가까울수록 좋다. '지니지수 = 0' 이라는것은 불확실한게 0 이라는 것이고 100%모두 잘 맞췄다(같은 특성을 가진, 같은 종류끼리 잘 모였다)라는 의미가 된다.  지니지수는 통계학의 복원추출 개념을 이용했기 때문에 제곱이 들어간다. 한 번만 측정하는 것은 우연히 그런일이 발생했을 수 있기 때문에 최소 두 번은 측정해봐야 정확하게 알 수 있다는 의미로 제곱을 해준다.  지니지수는 0.5가 최대치이며, 지니지수가 0.5라는것은 '반반 섞여있다'라는 의미이다. |
|  |  |