6/22 강의노트

확률적 경사하강법 : 딥러닝의 핵심

도표이(가) 표시된 사진

자동 생성된 설명

모델이 생성하는 예측값의 오차

실제값과 오차가 가장 작은 지점 (파란색)을 찾아내는 것이 모델 훈련의 궁극적인 목표

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

확률적 경사하강법 객체 생성

Loss : 손실함수 종류 지정(log는 로지스틱 함수를 의미)

Max\_iter : 에포크 횟수의 최댓값 지정

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

에포크 횟수가 적어서 ConvergenceWarning 경고로 모델이 수렴하지 않았다는 경고 확인

텍스트, 폰트, 스크린샷, 번호이(가) 표시된 사진

자동 생성된 설명

Partial\_fit : 에포크를 1회만 더 수행

학습과정에 진폭이 크기 때문에 에포크 횟수가 증가하더라도 그것이 과대적합 이전이라도 모델 성능이 감소할 수 있음

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Classes : 추출되는 테스트 데이터의 범주가 모든 데이터의 범주를 담지 않을 수 있으므로 전체 범주를 객체에 미리 입력하기 위해 생성하는 변수

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명

For (언더바) : 특별히 지정하는 변수 없이 그냥 n회 반복시킴

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

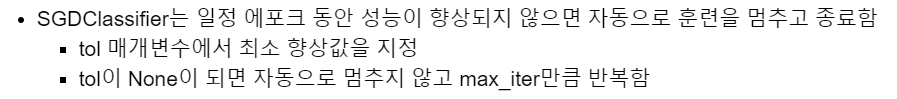
자동 생성된 설명

훈련데이터(파란색)과 테스트데이터(주황색)은 에포크가 진행될수록 성능이 증가, 일정 지점에서는 훈련데이터가 테스트데이터보다 성능이 낮은 구간도 있으며(과소적합) 안정화되면 훈련데이터의 성능이 더 좋아짐, 더욱 많은 에포크가 진행되면 과대적합이 되어 성능이 떨어질 것으로 예상

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명

Tol 매개변수 : 일정 에포크마다 훈련 로그를 확인하여 성능이 향상되지 않으면 훈련을 멈추는 기능을 지정, None을 입력하면 해당 기능을 사용하지 않음



텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

판다스의 원핫 인코딩(get\_dummies)보다 메모리 성능이 좋은 sklearn 원핫 인코더 사용 (난이도는 좀 더 어려움)

텍스트, 폰트, 스크린샷, 로고이(가) 표시된 사진

자동 생성된 설명

로지스틱 회귀나 선형회귀는 통계학 라이브러리로도 구현 가능

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Sm.OLS : 선형회귀 모델 생성 (종속변수, 독립변수 순서), sm.logic : 로지스틱 함수 모델 생성

Model = model.fit()으로 모델 훈련 진행

Model.summary로 통계적 수치 계산 결과 출력

R-Squared : 결정 계수

Adj. R-squared : 수정 결정 계수 (열이 많으면 사용)

F-statistic, Prob (F-statistic) : F 통계량에 대한 값, Prob(F-statistic)이 0.05 이하면 유의미한 결과를 나타냈다고 볼 수 있음 (실제로는 유의미하지 않으나 우연히 잘 예측할 확률, 낮을수록 좋음)

P>|t| 열 : 각각의 열이 연봉을 예측하는데에 유의미하게 사용되었는가 (0.05 이하일때 유의미함)