6/13 강의노트

텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

KNN회귀로 데이터 예측하기

텍스트, 폰트, 스크린샷, 번호이(가) 표시된 사진

자동 생성된 설명

KNN회귀 명령어셋은 독립변수가 2차원 데이터여야 함 (시리즈와 같은 1차원 배열은 한줄에 데이터가 몇 개가 들어가는지 인지하지 못해서 사용할 수 없음)

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

2차원으로 변환하여 사용

텍스트, 폰트, 영수증, 화이트이(가) 표시된 사진

자동 생성된 설명

모델 평가를 위한 결정계수 (타깃 = 예측하고자 하는 실제 값)

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

회귀 모델의 score은 결정계수를 나타냄’

텍스트, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

타깃과 예측값의 차이의 평균 (평균절대오차)으로도 평가 가능

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

KNN회귀모델의 한계점 : 데이터 범위를 벗어나는 값에 대한 예측에 약함

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

스크린샷, 텍스트, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

일정 이상의 길이에 대한 무게를 제대로 예측하지 못함

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명

선형회귀 모델

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

KNN회귀보다 올바른 값을 예측함

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

1차 그래프의 기울기와 y절편은 coef\_, intercept\_로 확인 가능

머신러닝을 통해 찾아낸 해당 값들은 모델 파라미터라고 부름 (model parameter)

최적의 모델 파라미터를 찾는 것이 머신러닝의 목표

선형 회귀 모델처럼 모델 파라미터가 있는 모델을 학습하는 것을 모델 기반 학습이라고 함 (KNN처럼 모델 파라미터가 없는 것은 사례 기반 학습이라고 함)

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

회귀선의 시각화

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

선형 모델의 평가, 점수 자체는 KNN보다 아래, 무게가 음수일수 없고, 데이터가 2차 곡선에 가까운 형태를 그리고 있어 2차이상의 식을 그려줄 필요가 있음

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

다항회귀 : 2차 이상의 곡선을 그리기 위하여 변수에 제곱한 열을 추가로 생성하여 모델 훈련에 사용

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

1차 직선보다 평가점수가 상승함

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

예측할 때에는 훈련시와 동일한 형태 [길이, 길이^2]로 입력

계수는 훈련값이 2종류이므로 리스트 형태로 2개의 값을 반환하였음

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명타이타닉 분류

Describe의 매개변수 include를 all로 지정하면 문자열에 대한 데이터도 출력됨 (최빈값을 포함)

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

분석에 사용할 데이터를 선택

텍스트, 스크린샷, 번호, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

범주형 데이터를 인식시키기 위하여 인코딩 실행

원핫인코딩으로 숫자형으로 변환, 다만 그 숫자가 데이터 분석에 숫자로서 사용되면 안됨

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

변환한 데이터를 원본 데이터에 붙여주기