

- 머신의 작동원리 · 어떤식으로 동작전행? = 머신을 작동하기 위함

2022.01.18 머신러닝

★ 패턴을 찾아서 값을 예측하는 형태. (정확한 값을 입력하는것이 중요)

✓ 지도학습 - Data: 정답을 틀다준다.

비지도학습 - Data를 주고 정답 모름 \Rightarrow pattern 찾고 단순히 4줄.

강화학습 - 시행착을 거쳐서 점차 Reward를 받아 좋은 형태로 보상.

★ 분류 vs 회귀

값의 범위 [값의 중간값이 의미가 있다 = 연속성이 5,500 (회귀)
고객의 이탈 : 1(yes) 0(no) 0.5=? (분류)

✓ target = Series의 형태로 가져온다.

\rightarrow data에서 x, y를 구분을 진행한다.

x_train	y_train
x_test	y_test

$x_train \Rightarrow y_train$ 으로 학습.

x_test는 평가, y_test는 (y-pred)
매번예측은 나온다.

Sklearn.model의 train_test_split.

$\langle x_train, x_test, y_train, y_test \rangle$

★ random_state = 섞는데, 섞는방법수따

\rightarrow 일정한 규칙으로 randomize

\rightarrow 실험 \rightarrow 같이 모델링의 성능을 진행하기위해, 동일한 평가를 위해서.

Scikit-Learn

① 불러오기 : import

② 선언하기 : train / test 데이터 분리. model

③ 학습하기 : model.fit

④ 예측하기 : model.predict

⑤ 평가하기 : 예측값과 실제값으로 평가

★ 시각화 (주로 범주형 데이터 시각화)

from sklearn.metrics import confusion_matrix

\Rightarrow 평가자료.

\rightarrow confusion_matrix(y_test, y_pred)

이후 Confusion-matrix를 seaborn을 이용하여 heatmap으로 시각화!

\Rightarrow

	0	1
0	a	b
1	c	d
\downarrow		
실제		

a: 0이라고 예측 \rightarrow 실제 0

b: 1이라고 예측 \rightarrow 실제 0

c: 0이라고 예측 \rightarrow 실제 1

d: 1이라고 예측 \rightarrow 실제 1

\Rightarrow 정확도

: accuracy score.

\rightarrow 자세한 내용은 내일

(recall, mse ...)