

2102. ML

1) 각 모델별 성능체크

✓ LinearRegression $\left\{ \begin{array}{l} \text{model.coef_} \\ \text{model.intercept_} \end{array} \right.$

`plt.barh(list(x), model.coef_)`

= 모델을 시각화 시켜, 어떤 값이 영향력이 있는지 눈으로 확인 가능

✓ Decisiontree \rightarrow `model.feature_importances_`

2) LogisticRegression. (분류에만 가능하다)

- 시그모이드함수를 이용한다.

(
 값이 커지면 1에 수렴
 값이 작아지면 0에 수렴

★ Recall 해당 결과가 실제로 그 결과일 확률.

K-Fold Cross Validation.

test 결과점수를 얼마나 맞을지 예측해보자.

(하나의 test data셋에서의 추정 → 객관적으로 이야기 하기 부족하다.)



☆ 의미

자체적으로 test 진행시, 높은 점수가 나왔다 = 가능성 ↑.

① 데이터셋을 random-state 지정X

→ 여러번 수행하고 accuracy 평균.

② 데이터 train-set을 등분짓고 (10등분) $\left(\begin{array}{l} 9\text{번 학습.} \\ 1\text{번 자체 test.} \end{array} \right)$ (10회)

⇒ 모든 데이터로 자체 평가진행. → 정확도 결과 (10개 추출).
average로 결과

학습데이터의 성능을 객관화 가능.