

22.01.19. ML

$$\left\{ \begin{array}{l} y: \text{실제값} \\ \hat{y}: \text{예측값} \\ \bar{y}: \text{평균값} \end{array} \right. \quad SST \left( \begin{array}{l} \text{SSE} \\ \text{SSR} \end{array} \right) \begin{array}{l} \bullet \text{ } y, \text{ 실제값} \\ \bullet \text{ } \hat{y}, \text{ 예측값} \\ \bullet \text{ } \bar{y}, \text{ 평균값} \end{array}$$

✓  $\sum (y - \bar{y})^2$  : **SST** (전체변이의 합, 최대한 이상에 성능은 보장되어야 함)

$\sum (\hat{y} - \bar{y})^2$  : **SSR** (전체변이에서 회귀선이 잡아낸 것)

$\sum (y - \hat{y})^2$  : **SSE** (전체변이에서 회귀선이 아직 잡아내지 못한 것)

$SST = SSR + SSE$

$R^2$  Score : 회귀모델의 성능,  $1 - \frac{E}{T} = \frac{R}{T}$

↓ 이득, 코딩방법

해당하는 결과 데이터 값이

무엇인지 확인 (이해 위주)

## 1. Confusion - matrix.

실제 \ 예측	0	1
0	20 TN	4 FP
1	6 FN	30 TP

정확도 (Accuracy) 얼마나 정확하게 예측?  $\frac{50}{60}$   
 정밀도 (precision)  $\frac{30}{34}$  (의정밀도) 예측값이 결과값에 부합?  
 재현율 (recall)  $\frac{30}{36}$  (의재현율) 결과값중 실제 결과값의 비율.  
 ↓  
 특이도 : 음성중, 음성이 정확할 확률 (0의 recall)

예측 : T/F 예측이 맞다/틀리다  
 실제 : N/P 예측이 0/1 이다.

## 2. F1 score (조화평균)

( precision  
recall

값이 작으면 조화평균이 작아진다.  
 주로 분자가 같고, 분모가 다른 경우 사용.

ex.

	0	1	pred
0	134	34	
1	43	57	
test			

$$\text{accuracy\_score} = \frac{134+57}{134+57+43+43} \cong \frac{191}{268} = 0.71268$$

$$\text{precision\_score} \left\{ \begin{array}{l} 0: \frac{134}{134+43} \cong 0.75706 \\ 1: \frac{57}{43+57} \cong 0.62637 \end{array} \right.$$

$$\text{recall} \left\{ \begin{array}{l} 0: \frac{134}{134+43} \cong 0.75706 \\ 1: \frac{57}{43+57} \cong 0.57 \end{array} \right.$$

1. accuracy - score, precision, recall  $\Rightarrow$  여러개의 변수가 존재. ( 전체 평균은 구해 1개의 값으로도 표현이 가능하나,  
→ average options 각각의 평균을 구하기 위해 average 값을 수정해줘야 한다.

1. macro = 라벨 별 각 합계 평균.

2. micro = 전체 평균

3. None = 라벨별 각 평균을 그대로 입력한다.

( Default : None.  
→ 기본적으로 "1"에 대해서만 표현. )

☆ 범주화 데이터의 경우, 가변수화 진행!

`Sns.heatmap( ..., format='g' )`

→ 정수의 형식으로 값표현.

✓ `plt.xticks( [0.5, 1.5, 2.5],  
[ 'Setosa', 'VersiColor', 'V'irginica' ] )`

와 같은 형식으로 범례표기가능.

(sum square error)

$$SSE : \sum (y - \hat{y})^2$$

$$MSE : \frac{SSE}{n} \Rightarrow RMSE = \sqrt{MSE}$$

$$MAE : \frac{\sum |y - \hat{y}|}{n}$$

mean absolute error

오차에 절대값.

MAPE

mean absolute percent error

$$R^2 \text{ Score} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$



✓ 회귀모델의 성능지표.

알고리즘의 원리

KNN → 몇개의 이웃.

⇒ 조금더 깊게 확인.

클래스의 수를 늘리기.

방명록.