

제12장. 분류 모형 선택과 평가



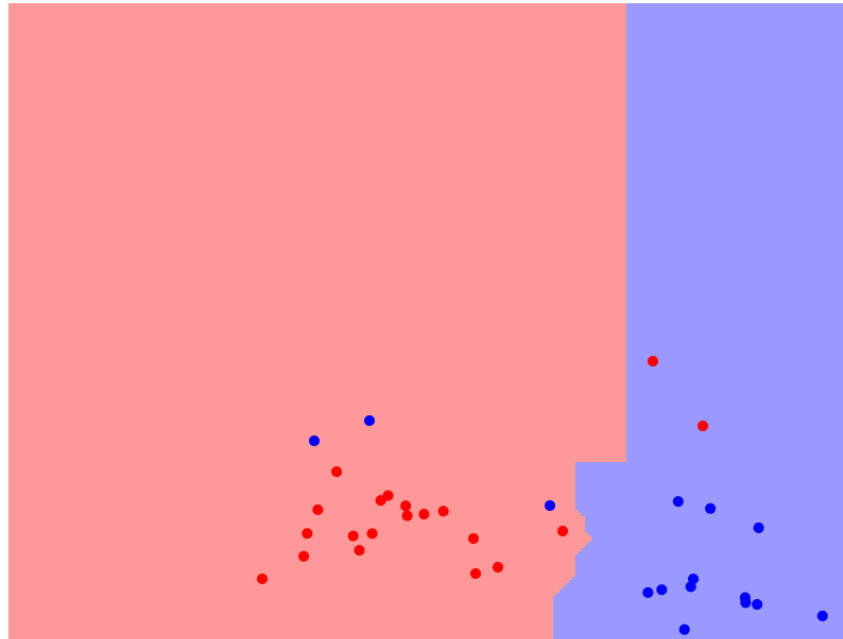
– Performance Metrics –

Study Point

- 모형의 선택과 평가에 대해 이해하고 실습한다.
- 예측 모형의 예측 값과 예측 확률의 의미를 이해한다.
- 혼동표는 예측 확률의 특정기준값에 따라 변한다는 것을 이해하고 실습한다.
- 분류 모형의 정확도(Accuracy), 검출율(Recall), 정밀도(Precision)을 이해하고 실습한다.

분류 예측 오류(KNN 알고리즘)

- 제 1종 오류 : 허위 양성 (False Positive : Negative → Positive), 실제로는 정상인데 비정상으로 판정
- 제 2종 오류 : 허위 음성 (False Negative : Positive → Negative), 실제로는 비정상인데 정상으로 판정



Metric

L1 L2

Num classes

2 3 4 5

Num Neighbors (K)

1 2 3 4 5 6 7

Num points

20 30 40 50 60





<http://vision.stanford.edu/teaching/cs231n-demos/knn/>

Accuracy, Recall, Precision

- 정확도(accuracy) : 예측 모형의 전체 정답율 = $(88+7)/100$
- 재현율(recall) : 실제 양성에서 양성을 올바르게 골라낸 비율 = $(7/10)$,
- 정밀도(precision) : 양성판정 중에서 결과의 정답 비율 = $(7/9)$,

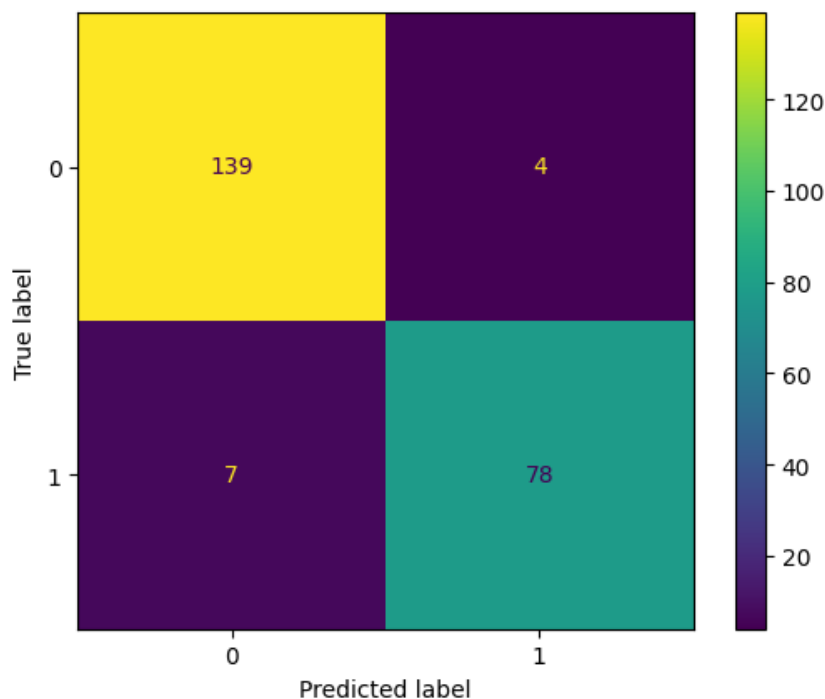
음성 재현율: 88/90

음성 정밀도: 88/91

		Negative (음성 판정) 통관	Positive (양성 판정) 검사	
Negative (음성)	적합약품	88  GOOD GOOD	2  GOOD BAD	90
	부적합약품	3 양성 클래스에 대한 잘못된 예측 비율  BAD GOOD	7  BAD BAD	10
		91	9	100

Accuracy, Recall, Precision

- 분류모델의 경우 소수범주 유형인 양성(Positive)을 분별하여 예측하는 것이 목적



$$N = TN(139) + FP(4) = 143$$

$$P = FN(7) + TP(78) = 85$$

정확도
(Accuracy)
 $(TN + TP) / (N + P)$

재현율, 참양성율
(Sensitivity, Recall)
 TP / P

정밀도
(Precision)
 $TP / (FP + TP)$

허위양성율
(False Positive Rate, FPR)
 FN / P

양성에 대한 Recall, Precision의 이해

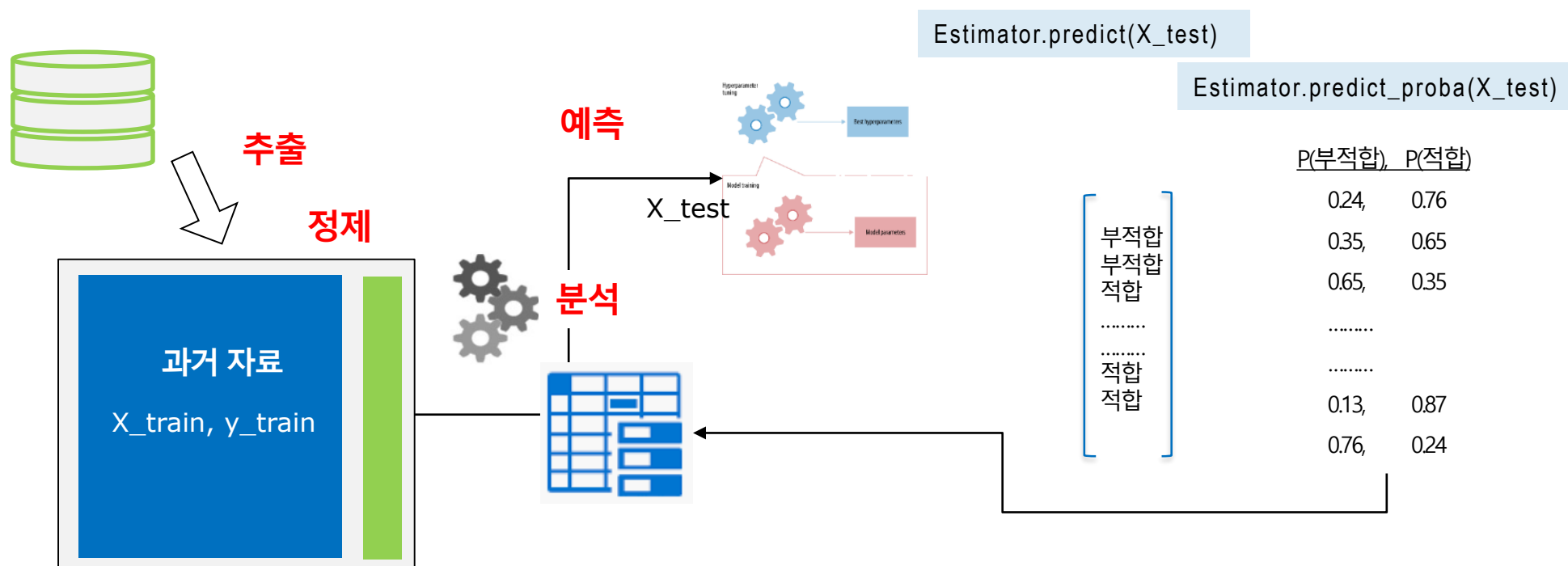
- 분류모델의 예측 오류는 허위음성과 허위양성의 2개의 종류가 있다
- 양치기 소년의 거짓말

허위 음성(False Negative) - 제2종 오류(Type II error) -	허위 양성(False Positive) - 제1종 오류(Type I error) -
<ul style="list-style-type: none">• 나쁜 사람을 좋은 사람으로 판단하여 결혼• 독버섯을 식용으로 판정하여 먹고 병원 응급실• 적군비행기를 새 떼로 판정하여 폭격을 당함• 살인자를 무죄로 판단하여 석방	<ul style="list-style-type: none">• 좋은 사람을 나쁜 사람으로 판단하여 이별• 식용을 독버섯으로 판정하여 기아상태• 세 떼를 적군비행기로 판정하여 대피하는 소동• 무죄인 사람을 살인자로 판단하여 감금



머신러닝 메커니즘(작동방식)

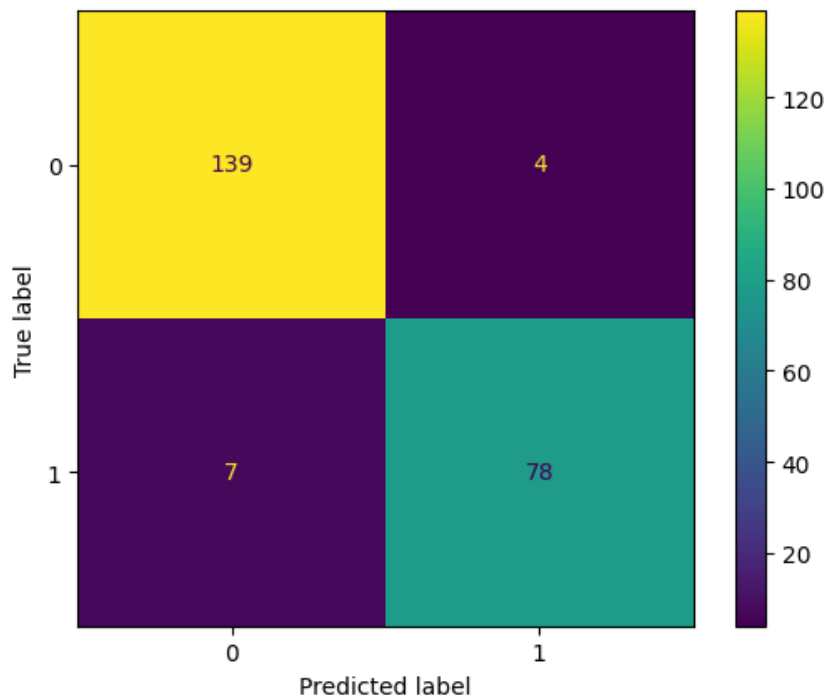
- 과거자료(X_{train} , y_{train})을 학습한 모형은 X_{test} 에 대한 분류 판정(적합, 부적합)과 분류 예측확률을 반환
 - scikit-learn의 알고리즘은 예측확률 0.5를 기준으로 분류판정을 나타낸다



분류 확률과 분류 판정

- Scikit-learn의 API는 예측 확률 0.5를 default 기준으로 분류 판정을 내린다.

```
y_predict = rf.predict(X_test)
y_proba = rf.predict_proba(X_test)
```



y_predict[:10]
(음성과 양성 판정)

```
array([
  1,
  0,
  0,
  0,
  0,
  0,
  1,
  0,
  1,
  1,
  1,
  1,
  1
])
```

y_proba[:10]
(음성과 양성 확률)

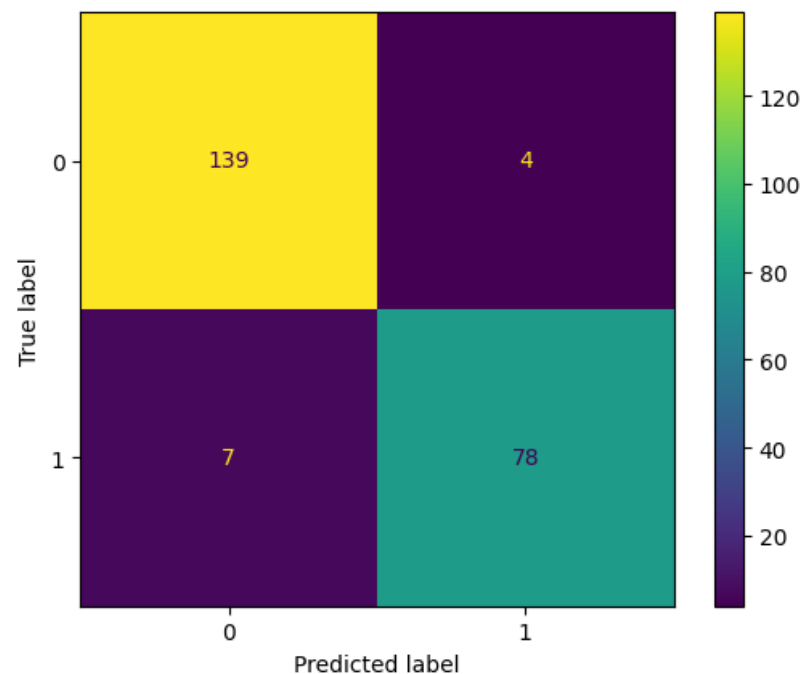
```
array([
  [0.02, 0.98],
  [0.92, 0.08],
  [0.99, 0.01],
  [0.98, 0.02],
  [0.99, 0.01],
  [0. , 1. ],
  [0.83, 0.17],
  [0.01, 0.99],
  [0. , 1. ],
  [0.33, 0.67]
])
```


Classification Report

- 양성과 음성에 대한 Classification report

```
from sklearn.metrics import classification_report
target_names = ['class 0', 'class 1']
print(classification_report(y_test, y_predict, target_names=target_names))
```

	precision	recall	f1-score	support
class 0	0.95	0.97	0.96	143
class 1	0.95	0.92	0.93	85
accuracy			0.95	228
macro avg	0.95	0.94	0.95	228
weighted avg	0.95	0.95	0.95	228



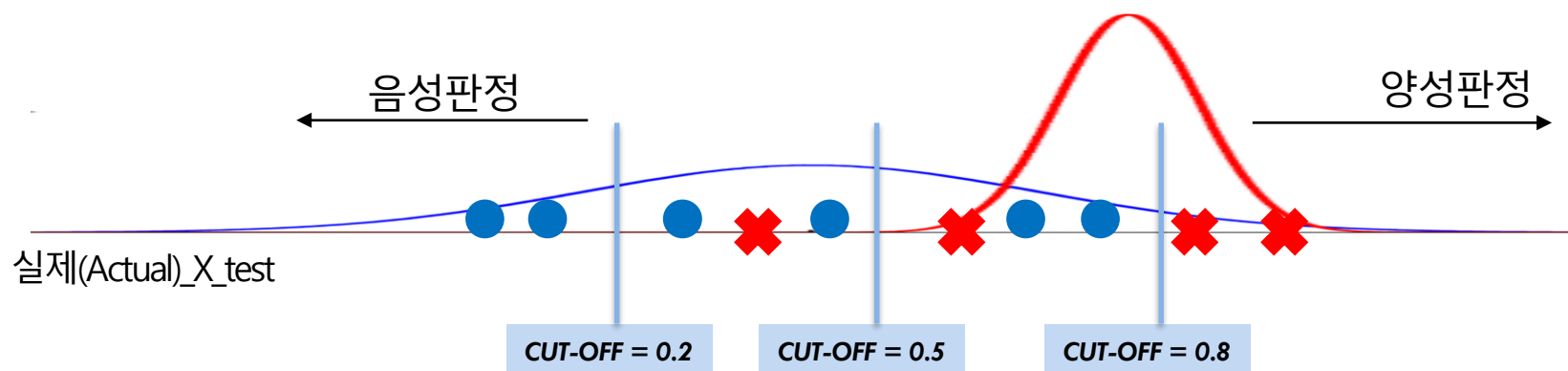
Classification Report 산술식

	A	B	C	D	E	F
1			Predicted label			
2			0	1		
3	True label	0	139	4		
4		1	7	78		
5						
6			precision	recall	f1-score	support
7		class 0	0.95	0.97	0.96	143
8		class 1	0.95	0.92	0.93	85
9						
10		accuracy			0.95	228
11		macro avg	0.95	0.94	0.95	
12		weighted avg	0.95	0.95	0.95	

	A	B	C	D	E	F
1			Predicted label			
2			0	1		
3	True label	0	139	4		
4		1	7	78		
5						
6			precision	recall	f1-score	support
7		class 0	=C3/(C3+C4)	=C3/F7	=2*(C7*D7)/(C7+D7)	=C3+D3
8		class 1	=D4/(D4+D3)	=D4/F8	=2*(C8*D8)/(C8+D8)	=C4+D4
9						
10		accuracy			=(C3+D4)/F10	=SUM(F7:F9)
11		macro avg	=AVERAGE(C7:C8)	=AVERAGE(D7:D8)	=AVERAGE(E7:E8)	
12		weighted avg	=C7*F7/F10 + C8*F8/F10	=D7*F7/F10+D8*F8/F10	=E7*F7/F10+E8*F8/F10	

Cut-off 값

- 양성과 음성 판정기준이 되는 예측 확률을 '0.1', '0.5', '0.8' 등 연구자의 목적에 따라 분류 판정을 내릴 수 있다.



Confusion Matrix	CUT-OFF = 0.2		CUT-OFF = 0.5		CUT-OFF = 0.8		Support
	N	P	N	P	N	P	
	N	2	4	4	2	6	
	P	0	4	1	2	2	4
		2	8	5	8	2	10

양성에
대한
Recall,
Precision

판정기준	0.2	0.5	0.8
Recall	4/4=1.00	3/4=0.75	2/4=0.50
Precision	4/8=0.50	3/5=0.60	2/2=1.00

의사결정임계치

- 카드 사기, 부도업체 적발, 불량 제품 예측 등에는 허위음성(FN)를 줄이는 것이 중요하다.
- Recall과 Precision은 반비례 관계이다.
- 양성판정을 엄격하게 하기 위해 Cut-off 값을 증가시키면 참 양성 개수(TP)는 감소

