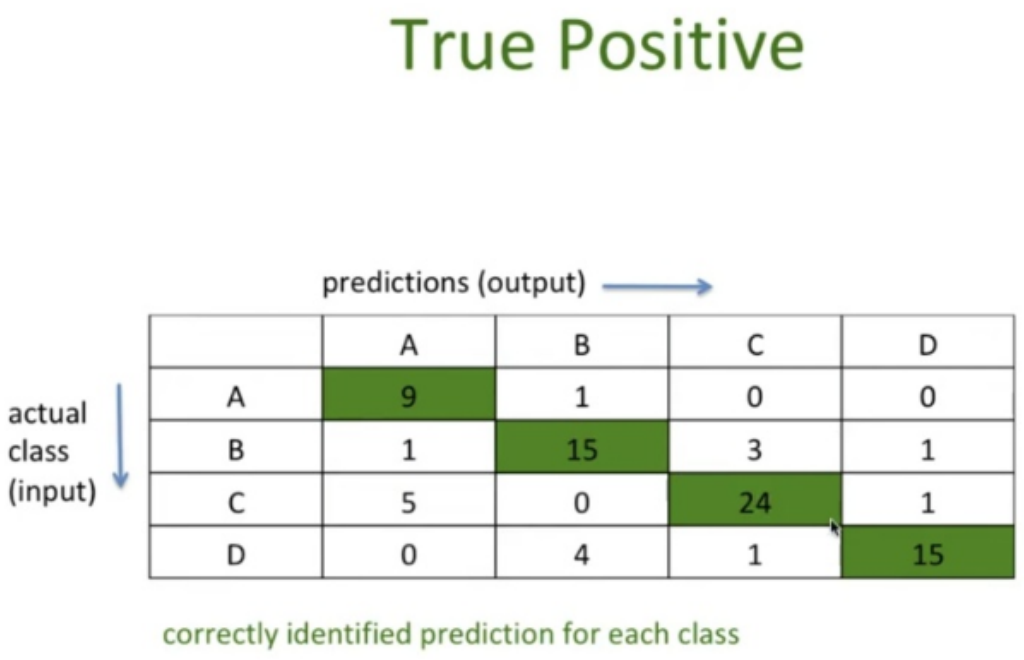
분류성능평가지표

분류성능평가지표

TP (True Positive) - 실제값과 예측값이 모두 True인 경우 (정답을 정답이라 본 경우)



TN (True Negative) - 실제값과 예측값이 모두 False인 경우 (오답을 오답이라 본 경우)

True Negative for A

predictions (output) →

actual class (input) ↓

	A	B	C	D
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

correctly rejected prediction for certain class (A)

FP (False Positive) - 실제로 False이지만 예측은 True인 경우 (오답을 정답이라 본 경우)

False Positive for A

predictions (output) →

actual class (input) ↓

	A	B	C	D
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

incorrectly identified predictions for certain class (A)

FN(False Negative) - 실제로 True이지만 예측은 False인 경우 (정답을 오답으로 본 경우)

False Negative for A

predictions (output) →

actual class (input) ↓

	A	B	C	D
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

incorrectly rejected for certain class (A)

Accuracy_score (정확도) - TP / 전체

Accuracy

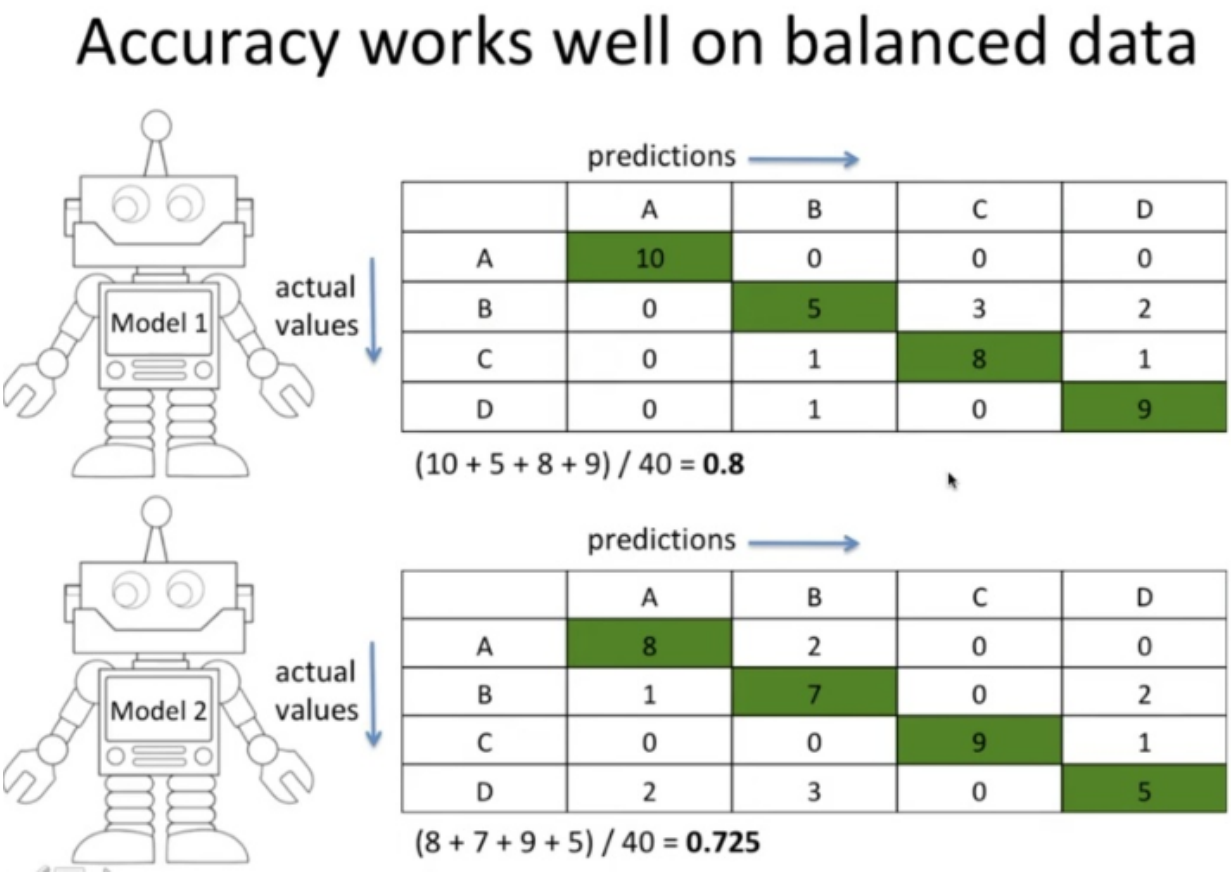
predictions (output) →

actual class (input) ↓

	A	B	C	D
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

correctly identified prediction for each class / total dataset
9 + 15 + 24 + 15 / 80
accuracy = 0.78

정확도는 데이터의 분포가 고루 퍼져있을 때 유용하다.



F1 - Precision*recall / Precision + reall

정확도와 재현율의 조화 평균

F1은 데이터의 분포가 고르지 못한 상태의 정확성을 검증할 때 유용하다.

F1 score is good metric when data is imbalanced

Given a class, will the classifier detect it ? (recall)

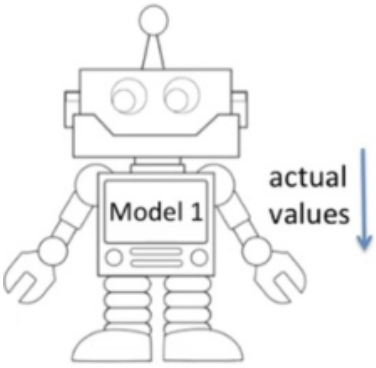
	A	B	C	D
A	100	80	10	10
B	0	9	0	1
C	0	1	8	1
D	0	1	0	9

Given a class prediction from the classifier,
how likely is it to be correct? (precision)

F1 Score is harmonic mean of recall and precision

Precision(정밀도) = TP / (TP+FP) - 예측이 True인 것 중 실제로 True인 비율

Precision of Model 1 (macro average)



predictions →

actual values ↓

	A	B	C	D
A	100	80	10	10
B	0	9	0	1
C	0	1	8	1
D	0	1	0	9
	TP: 100 FP: 0	TP: 9 FP: 82	TP: 8 FP: 10	TP: 9 FP: 12

Precision = TP / (TP + FP) P(A) = 1 P(B) = 9/91 P(C) = 8/18 P(D) = 9/21

average precision = P(A) + P(B) + P(C) + P(D) / 4 = 0.492

the number of classes

Recall(재현율) = TP / (TP + FN) - 실제값이 True인 것 중 예측이 적중한 비율

Recall of Model 1 (macro average)

predictions →

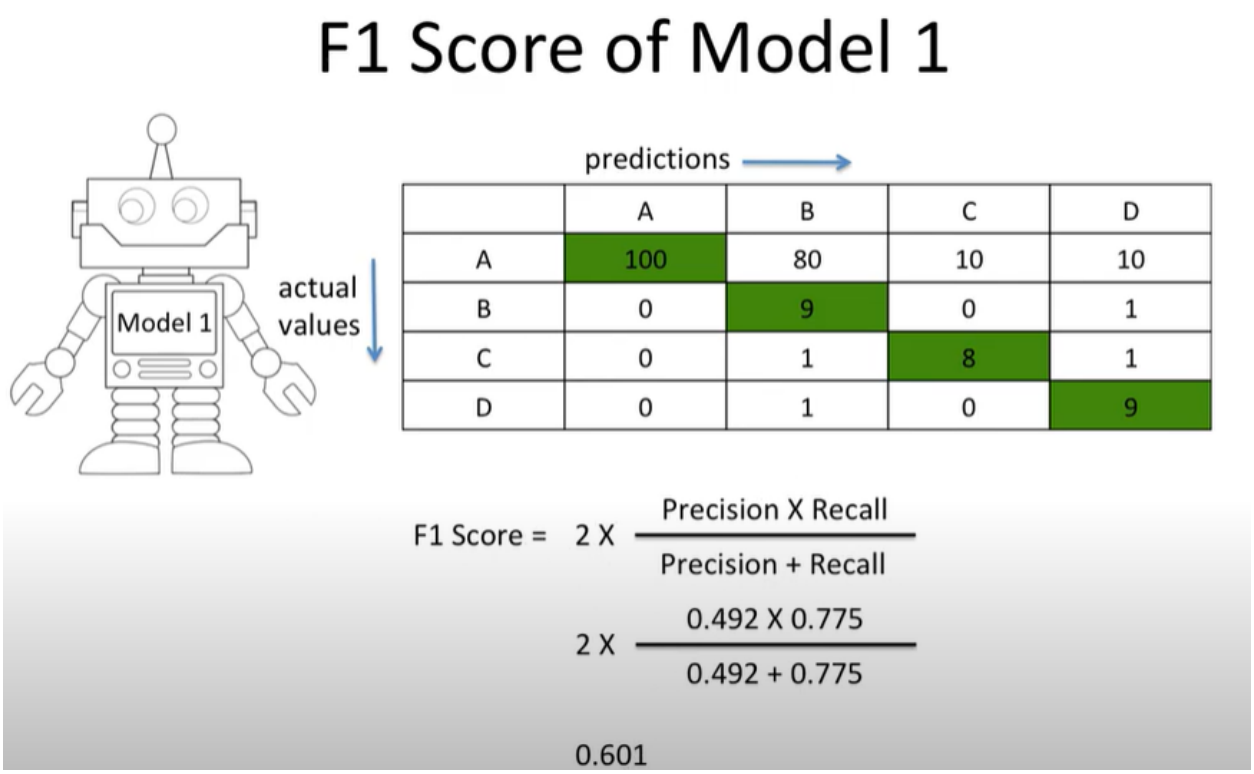
	A	B	C	D			
A	100	80	10	10	TP: 100,	FN: 100	R(A) = 100 / 200
B	0	9	0	1	TP: 9,	FN: 1	R(B) = 9/10
C	0	1	8	1	TP: 8,	FN: 2	R(C) = 8/10
D	0	1	0	9	TP: 9,	FN: 1	R(D) = 9/10

Recall = TP / (TP + FN)

average recall = R(A) + R(B) + R(C) + R(D) / 4 = 0.775

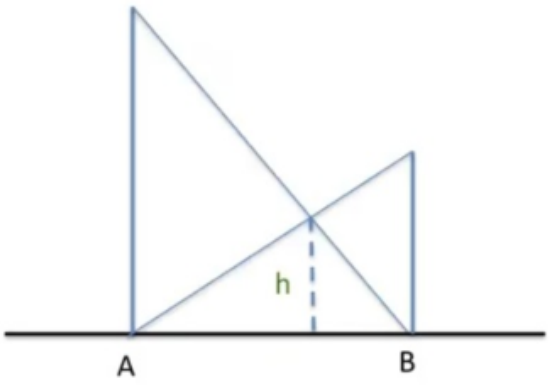
the number of classes

F1_Score의 원리



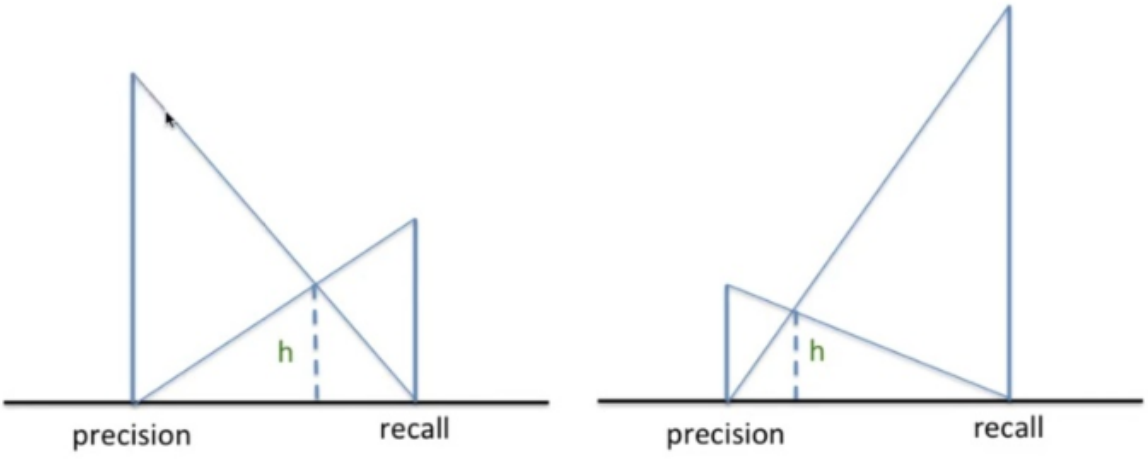
Harmonic Mean - F1의 원리로 조화 평균을 의미

Harmonic Mean



h is half the harmonic mean

Harmonic Mean punishes extreme value more



h is half the harmonic mean

$$\text{F1 Score} = 2 \times \frac{\text{Precision X Recall}}{\text{Precision + Recall}}$$

분류성능평가지표 한눈에 보기 with Python

```
from sklearn.metrics import classification_report
print(classification_report(val_y, result))

                 precision    recall  f1-score   support

            0       1.00      1.00      1.00        13
            1       0.92      1.00      0.96        11
            2       1.00      0.93      0.96        14

    accuracy                           0.97        38
   macro avg       0.97      0.98      0.97        38
weighted avg       0.98      0.97      0.97        38
```

Average

```
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score
print('정확도 {}'.format(accuracy_score(val_y, result)))
print('정밀도 {}'.format(precision_score(val_y, result, average='micro')))
print('재현율 {}'.format(recall_score(val_y, result, average='micro')))
print('F1 {}'.format(f1_score(val_y, result, average='micro')))

average = 'None' 라벨 별 각 평균
average = 'micro' 전체 평균
average = 'macro' 라벨 별 각 평균의 합
average = 'weighted_avg' 라벨별 가중치(개수)를 부여한 각 평균
```