

평가 지표(Metric)

목차

4-1 회귀에서의 평가 지표

4-2 분류에서의 평가 지표

4-1 회귀에서의 평가 지표

MAE(mean absolute error, MAE) : 평균 절대 오차

MSE(Mean Square Error; MSE) : 평균 제곱 오차

RMSE(Root Mean Square error, RMSE) : 평균 제곱근 오차(RMSE)

RMSLE(Root Mean Squared Logarithmic Error, RMSE)

MAPE(Mean Absolute Percentage Error : MAPE) : 평균 절대 백분 오차 비율

4-1 예측 오차

\hat{y}_i : 모델로부터 생성된 예측값

y_i : 실제값

$y_i - \hat{y}_i$ 실제값과 예측값의 차이

4-1 예측 오차

▶ MAE(Mean absolute Error; MAE) : 평균 절대 오차

* 예측 에러(Prediction Error) = 실제값 - 예측값

* 절대 오차(Absolute Error) -> |예측 오차|

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|)$$

\hat{y}_i : 모델로 예측한 값

y_i : 실제값

n : 데이터 개수

* MAE는 오차이기에 0에 가까우면 좋다.

4-1 예측 오차

▶ **MSE(Mean Squared Error; MSE) : 평균 제곱 오차**

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

\hat{y}_i : 모델로 예측한 값

y_i : 실제값

n : 데이터 개수

* MSE는 오차이기에 0에 가까우면 좋다.

4-1 예측 오차

▶ RMSE(Root Mean Square error, RMSE) : 평균 제곱근 오차

RMSE는 평균제곱오차(MSE)의 제곱근 값으로 다음과 같이 정의한다.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

\hat{y}_i : 모델로 예측한 값

y_i : 실제값

n : 데이터 개수

* MSE로 예측 오차를 평가할 때, 그 수치가 커지는 것을 제곱근을 취함으로써 보정한 값이다.

4-1 예측 오차

▶ MAPE (Mean Absolute Percentage Error, MAPE)

- 평균 절대 백분 오차 비율 (Mean Absolute Percentage Error : MAPE)
- MAPE는 퍼센트 값을 가진다. 0에 가까울수록 회귀 모델의 성능이 좋음.
- MAPE는 실제 target (종속변수) 값 대비 예측오차 비율의 절대값들을 평균한 값으로 실제 데이터에서 오차가 어느 정도의 비율로 발생했는지 확인한다.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

\hat{y}_i : 모델로 예측한 값

y_i : 실제값

n : 데이터 개수

* MAPE는 0~100% 사이의 값을 가져 이해하기 쉬워 성능 비교 해석이 가능

4-2 분류 문제에서는 어떤 평가지표를 사용할까?

4-2 Metric

▶ 혼동 행렬(confusion matrix) or 오차 행렬

- Confusion matrix는 분류 모델의 성능을 알고자 할 때 많이 사용.
- 알고리즘 성능을 평가할 때, 평가하는 지표로 많이 사용.

4-2 Metric

▶ 이진 분류 모델을 위한 혼동 행렬

- 파란색 부분은 정답을 맞춘 경우이고, 붉은색 부분은 틀린 부분

	Actual(실제) = Y	Actual(실제) = N
Predict(예측)=Y	True Positive(TP)	False Positive(FP)
Predict(예측)=N	False Negative(FN)	True Negative(TN)

4-2 Metric

▶ 혼동 행렬(confusion matrix)

(2) 'P', 'N'에서 'P'는 'Positive', 'N'는 'Negative'

	Actual(실제) = Y	Actual(실제) = N
Predict(예측)=Y	True Positive(TP)	False Positive(FP)
Predict(예측)=N	False Negative(FN)	True Negative(TN)

(1) 앞의 'True'는 실제와 예측이 일치. 'False'는 예측이 일치하지 않음.

(2) 뒤의 Positive, Negative는 예측을 무엇으로 했는가? Y(Positive), N(Negative)

4-2 Metric

▶ 혼동 행렬(confusion matrix)

	Actual(실제) = Y	Actual(실제) = N
Predict(예측)=Y	True Positive(TP)	False Positive(FP)
Predict(예측)=N	False Negative(FN)	True Negative(TN)

(1) TP의 경우, 예측 성공(True). 실제 Y(1)의 값을 1(Positive)로 예측

(2) TN의 경우, 예측 성공(True) 실제 N(0)의 값을 0(Negative)으로 예측

4-2 Metric

▶ 혼동 행렬(confusion matrix)

	Actual(실제) = Y	Actual(실제) = N
Predict(예측)=Y	True Positive(TP)	False Positive(FP)
Predict(예측)=N	False Negative(FN)	True Negative(TN)

(1) FP의 경우 예측 실패(False). 실제 N(0)의 값을 1(Positive)로 예측

(2) FN의 경우 예측 실패(False). 실제 Y(1)의 값을 0(Negative)로 예측

4-2 Metric

▶ 정확도(Accuracy)

전체 샘플 데이터 중에서 예측한 값이 얼마나 잘 맞았을까?

	Actual=Y	Actual=N
Predict=Y	TP	FP
Predict=N	FN	TN

정확도(Accuracy) =

$$\frac{TP + TN}{TP + FP + FN + TN}$$

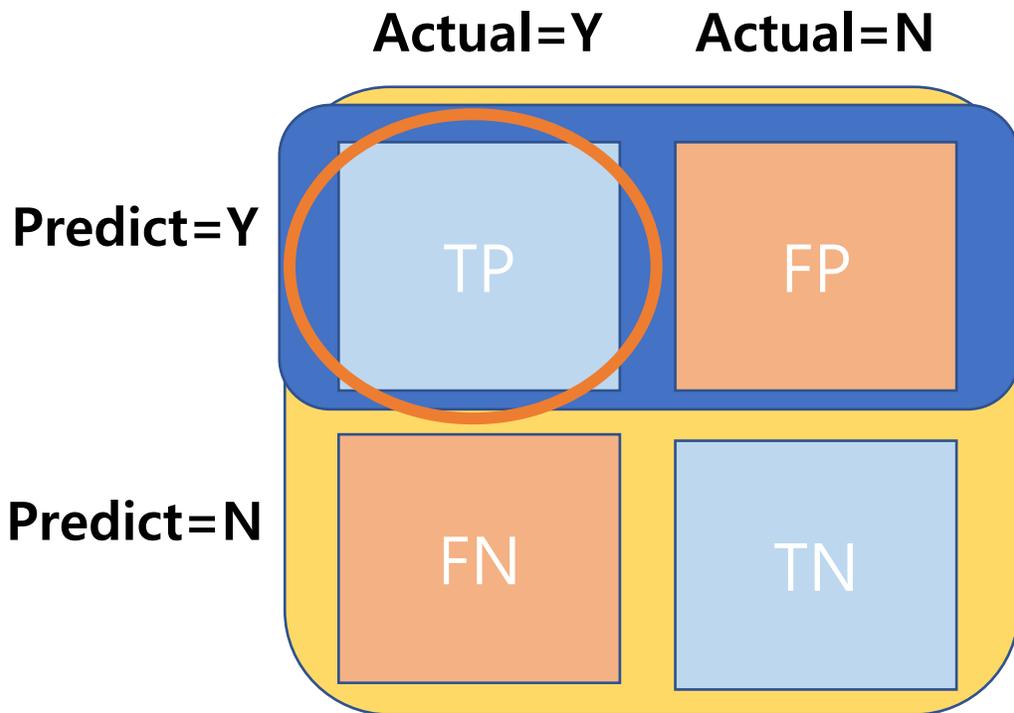
TP	FP
FN	TN

TP	FP
FN	TN

4-2 Metric

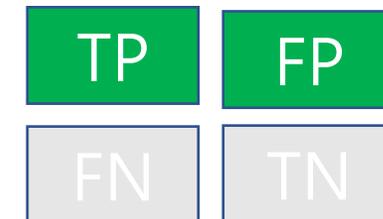
▶ 정밀도(Precision)

예측 Y중에 실제 Y를 정확하게 한 확률



정밀도(Precision) =

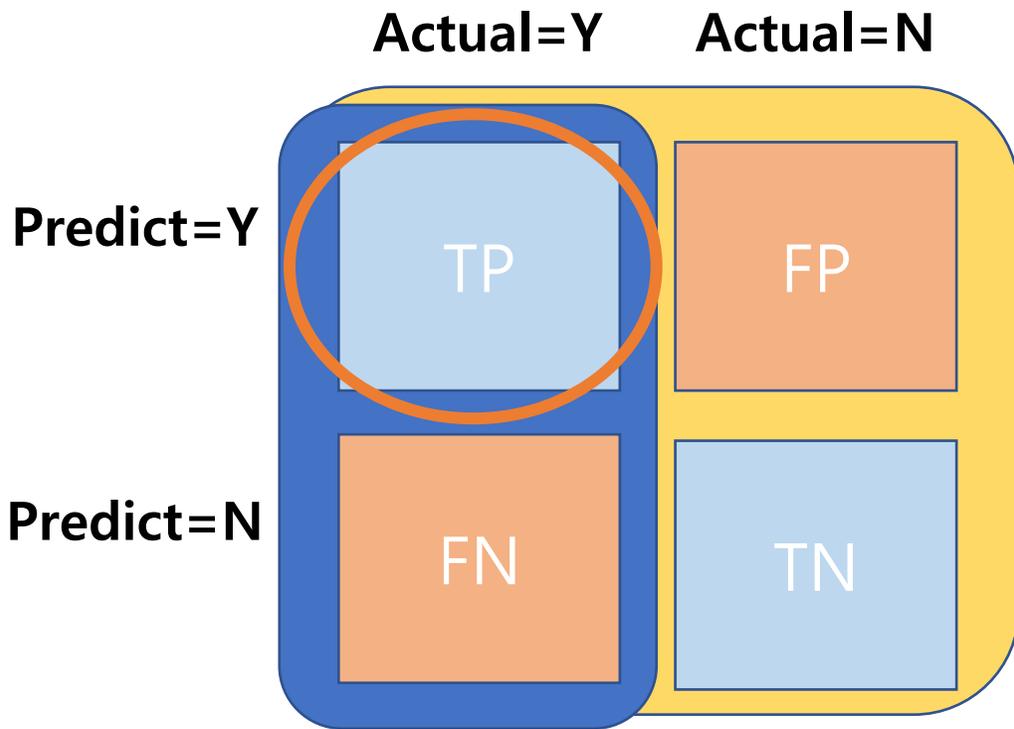
$$\frac{TP}{TP + FP}$$



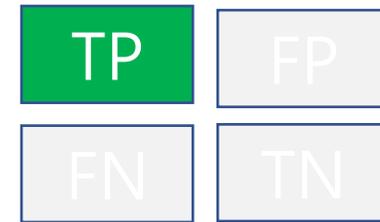
4-2 Metric

▶ 민감도(Sensitivity, Recall, TPrate)

실제 Y값 중에서 예측을 정확하게 Y로 한 확률



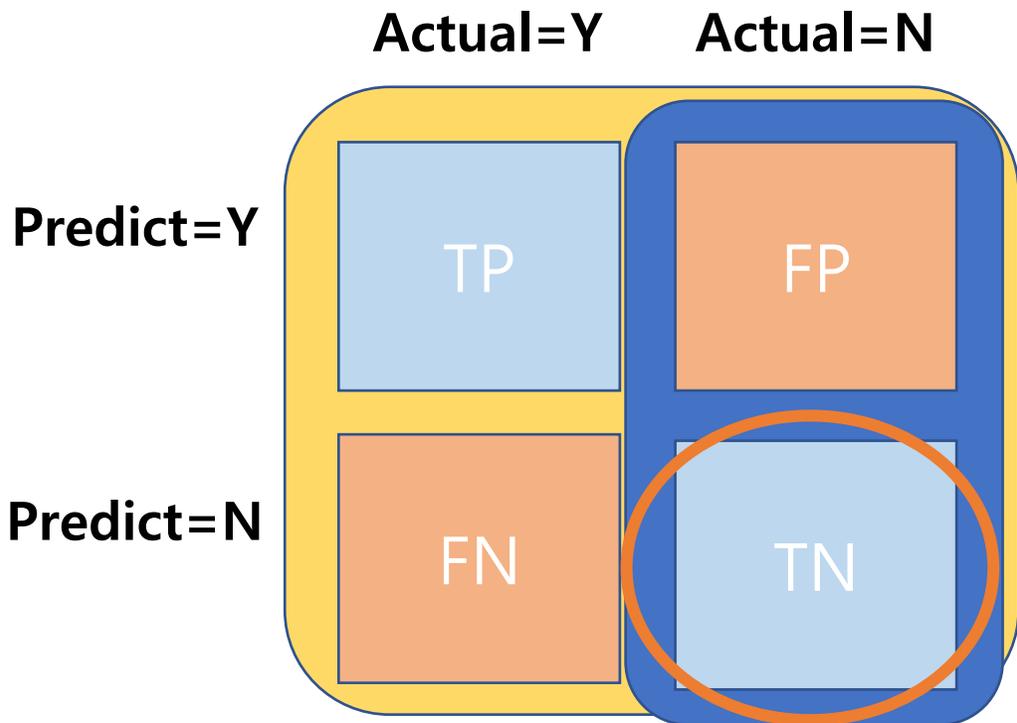
$$\text{민감도(Sensitivity)} = \frac{TP}{TP + FN}$$



4-2 Metric

▶ 특이도(Specificity)

실제 N(거짓)인 것 중에서 정확하게 N(거짓)을 예측한 확률



TP	FP
FN	TN

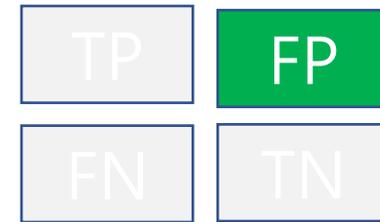
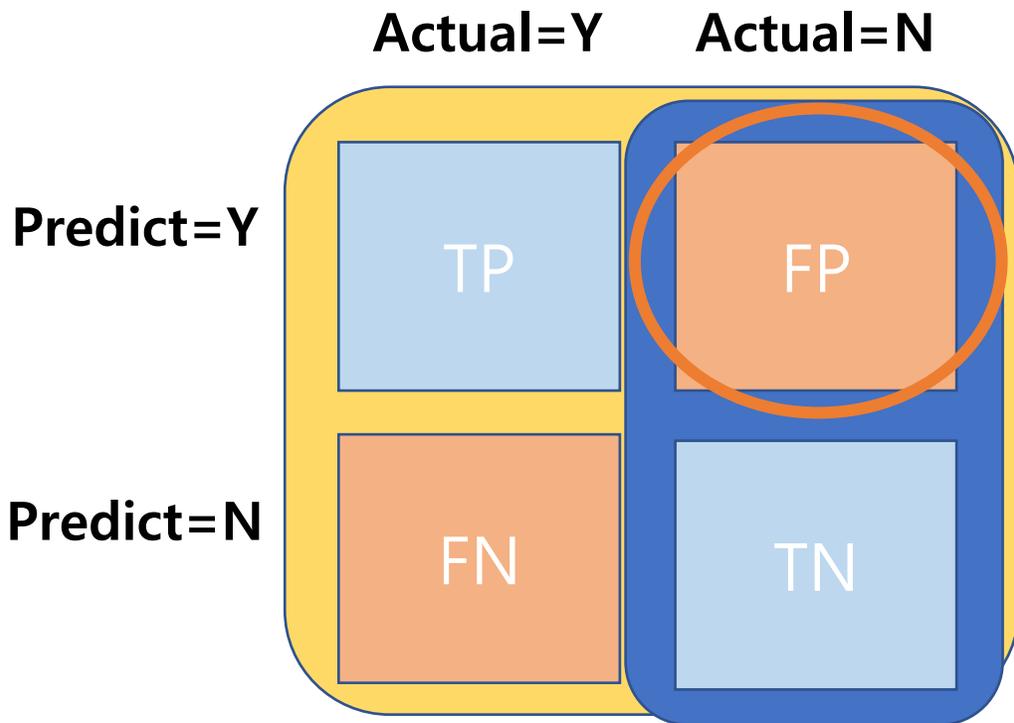
$$\text{특이도 (Specificity)} = \frac{TN}{FP + TN}$$

TP	FP
FN	TN

4-2 Metric

▶ Fprate – 틀린 케이스

실제로 N(거짓)중에 예측이 틀린 경우(N인데 예측을 Y) 한 경우



$$\text{Fprate} = \frac{FP}{FP + TN}$$

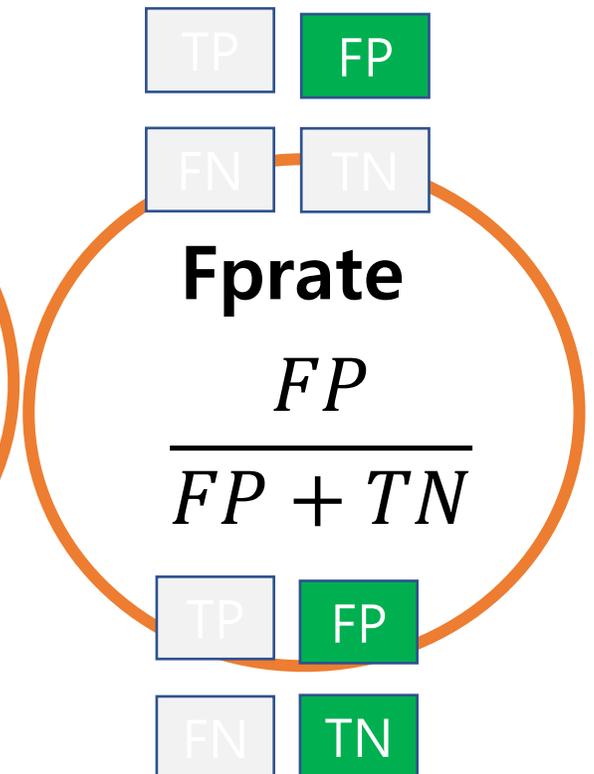
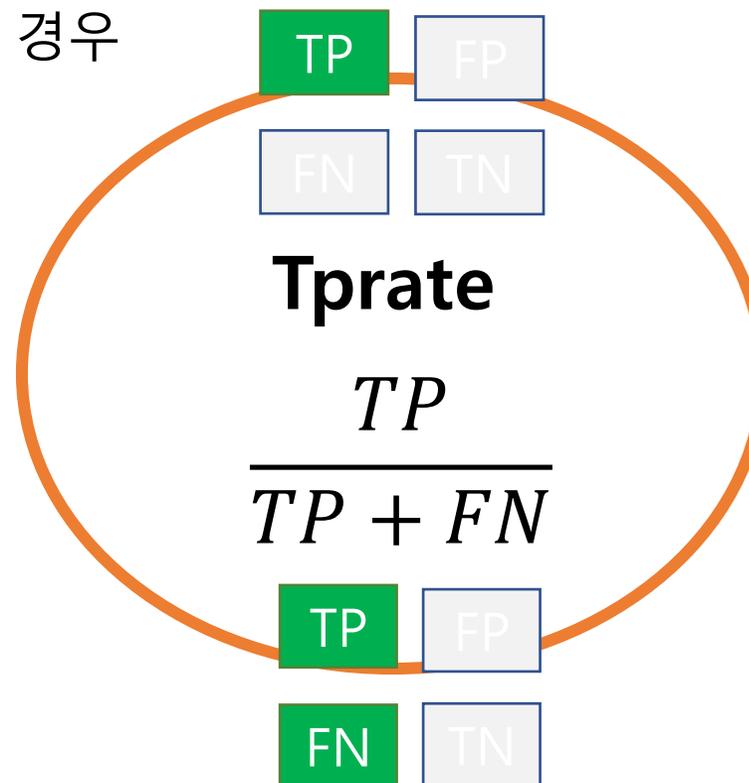
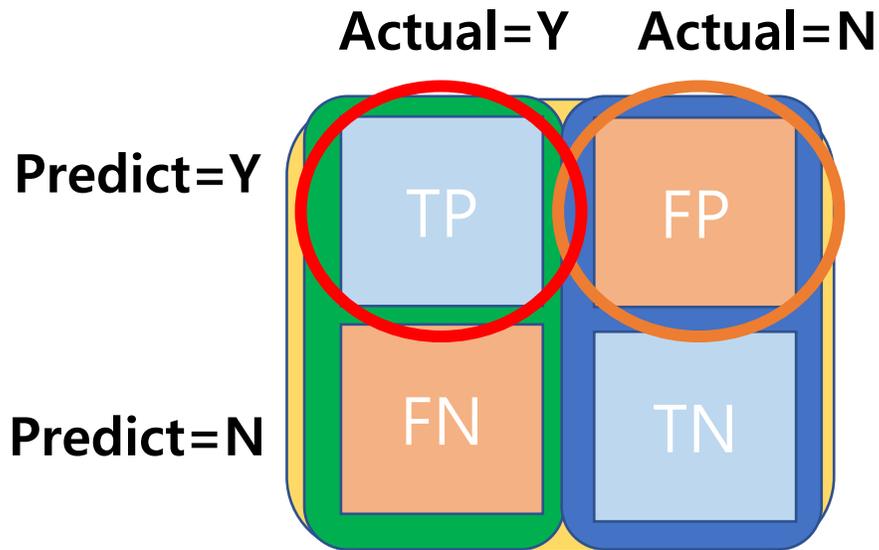


4-2 Metric

▶ 민감도(Sensitivity, Recall, TPrate)와 FPrate

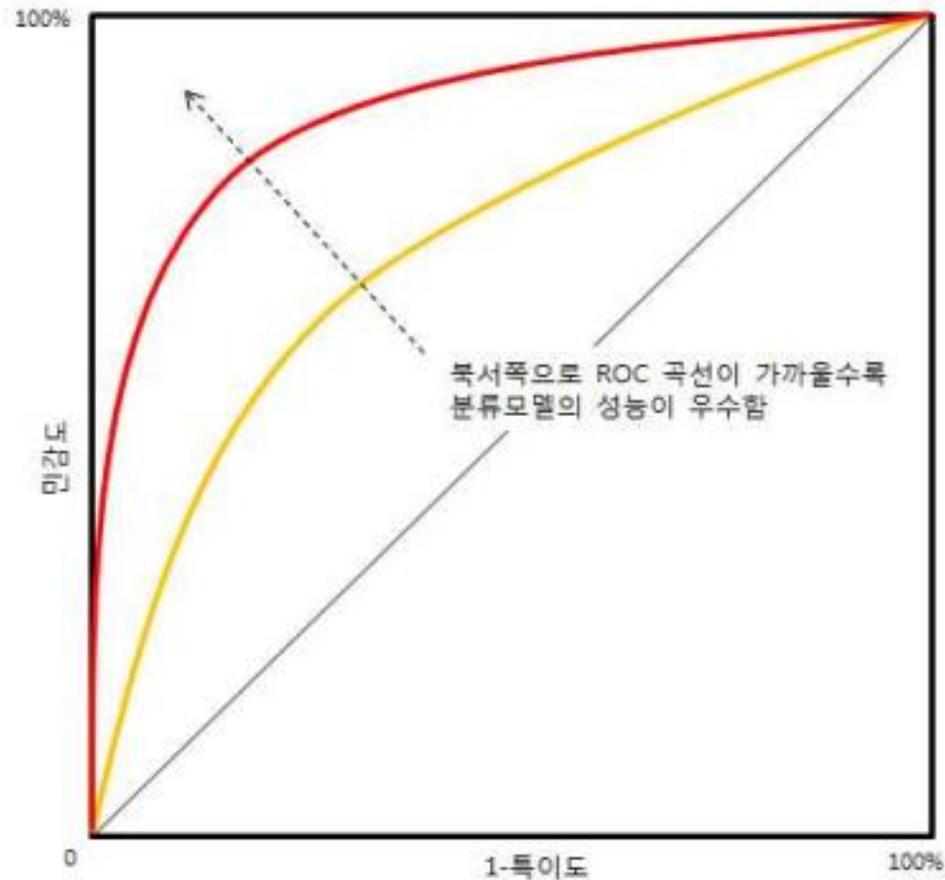
민감도 : 실제로 Yes인데 예측이 Yes로 된 경우

Fprate : 실제로 No인데 예측이 Yes로 된 경우



4-2 Metric

▶ ROC Curve

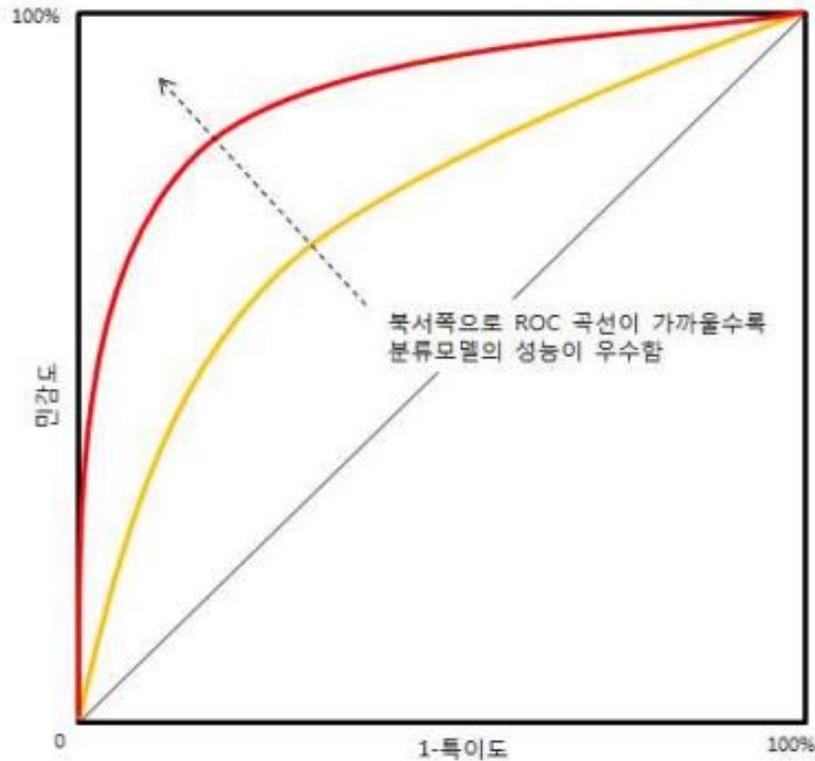


- 분류 문제와 관련하여 AUC-ROC곡선을 성능 측정용으로 사용할 수 있다.
- AUC(곡선 아래 영역)을 의미함.
- ROC는 확률 곡선을 의미함.

[그림 3-2] ROC 곡선

4-2 Metric

▶ ROC Curve



[그림 3-2] ROC 곡선

- AUC가 높을수록 모델을 0을 0으로 1을 1로 예측하는 것이 좋음.

4-2 Metric

▶ f-score

- f-score는 의미상으로는 정밀도(Precision)과 민감도(Recall)에 대한 평균
- 가중치를 주는 평균이라 생각함.

$$f_{\beta} = \frac{(1 + \beta^2)(\text{정밀도} \cdot \text{민감도})}{(\beta^2 \cdot \text{정밀도} + \text{민감도})}$$

4-2 Metric

▶ F1-score

- F-score의 일반식에서 $\beta = 1$ 인 경우를 f1-score라고 한다.

$$f_1 = \left(\frac{recall^{-1} + precision^{-1}}{2} \right)^{-1} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

위의 내용은 결국 precision과 recall의 Harmonic mean(=Harmonic average)가 된다.

4-2 Metric

▶ 조화평균(Harmonic Mean)

조화 평균이란,
n개의 양수에 대하여 그 역수들을 산술 평균한 것의 역수를 말함.

$$f_1 = \left(\frac{recall^{-1} + precision^{-1}}{2} \right)^{-1} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$