

데이터 마이닝 모델 평가

목차

4-1 예측 오차

4-2 Metric

4-1 예측 오차

\hat{y}_i : 예측 모델을 적용하여 예측한 값

y_i : 검증 데이터 세트의 실제 값

$y_i - \hat{y}_i$ 를 이용하여 산출하는 척도

MAD(Mean Absolute Deviation, MAD) : 평균 절대 편차

MAE(Mean absolute error; MSE) : 평균 절대 오차

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

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

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

4-1 예측 오차

MAE(Mean absolute error; MAE) : 평균 절대 오차

MAE는 예측 오차의 절대값들 평균을 나타내는 방법이다.

$$\text{MAE} = \frac{1}{n} \sum_{y_i \in L} |y_i - \hat{y}_i|$$

4-1 예측 오차

MSE(Mean squared error; MSE) : 평균 제곱 오차

MSE는 실제값과 예측값의 차의 제곱을 관측치 개수로 나눈 값

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

4-1 예측 오차

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

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

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

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 Absoulte Percentage Error : MAPE)

MAPE는 실제 종속변수 값 대비 예측오차 비율의 절대값들을 평균한 값으로 실제 데이터에서 오차가 어느 정도의 비율로 발생했는지 확인한다.

$$\text{MAPE} = \frac{1}{n} \sqrt{\sum_{y_i \in L} \frac{|y_i - \hat{y}_i|}{y_i}}$$

4-2 Metric

혼동 행렬(confusion matrix)

파란색 부분은 정답을 맞춘 경우이고, 붉은색 부분은 오류가 생성된 부분

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

4-2 Metric

정확도(Accuracy)

예측한 값이 얼마나 잘 맞았을까?

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

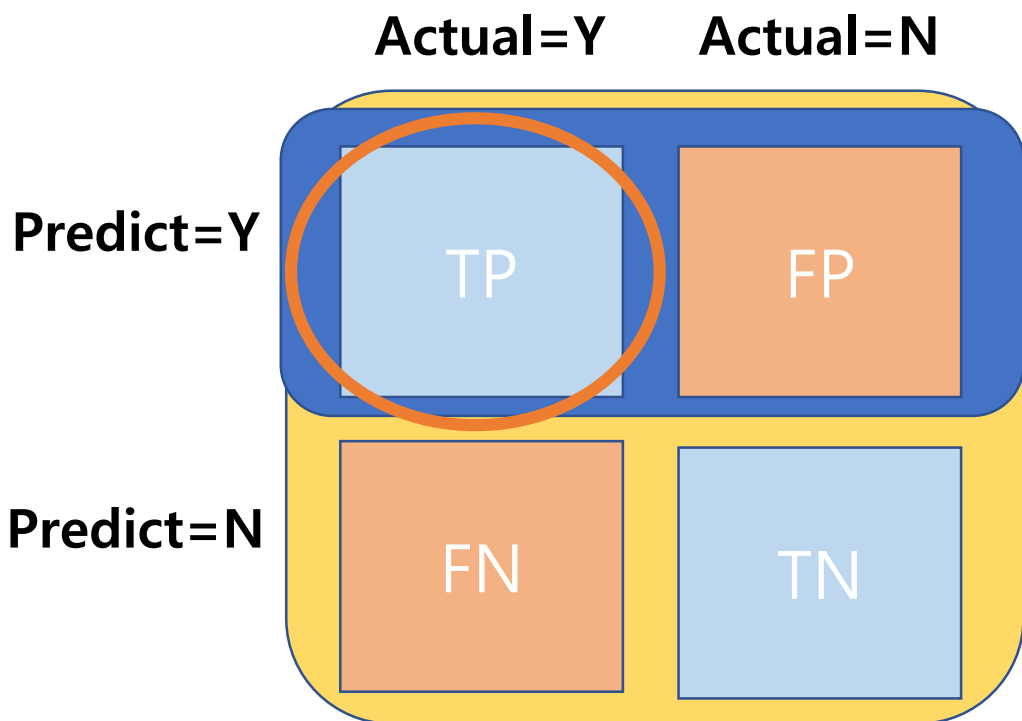
정확도(Accuracy) =

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

4-2 Metric

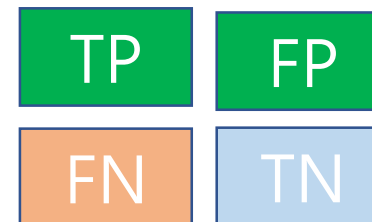
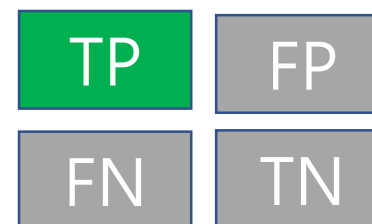
정밀도(Precision)

실제 Y, N의 값 중에서 예측을 Y로 한 확률



정밀도(Precision) =

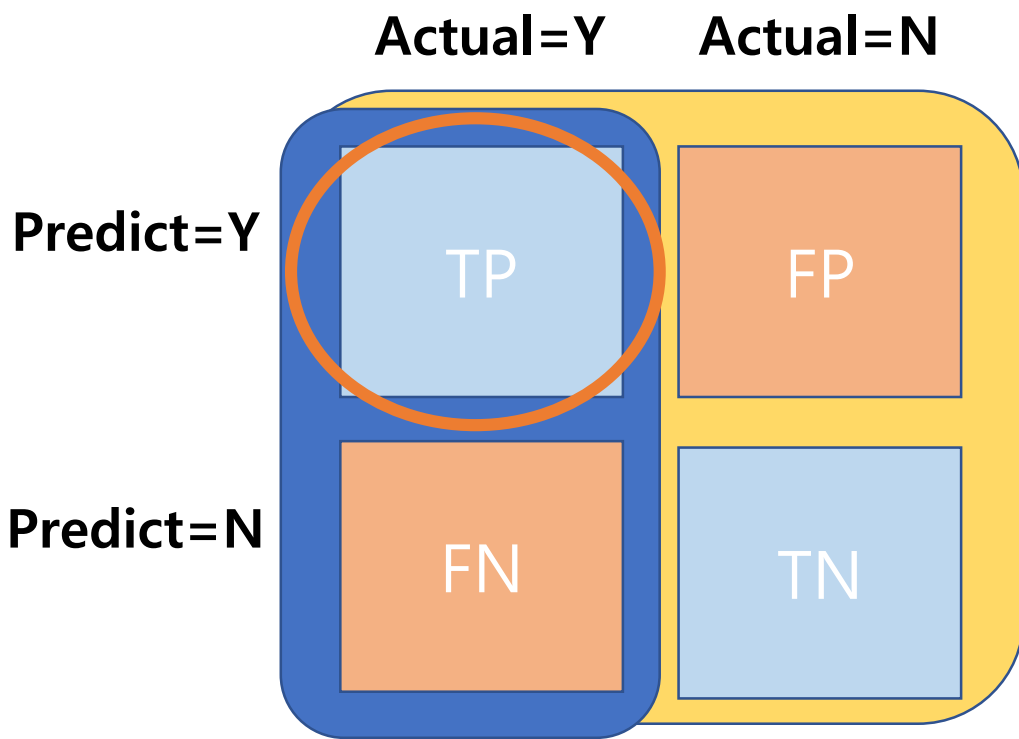
$$\frac{TP}{TP + FP}$$



4-2 Metric

민감도(Sensitivity, Recall, TPrate)

실제 Yes값 중에서 예측을 Yes로 한 확률



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

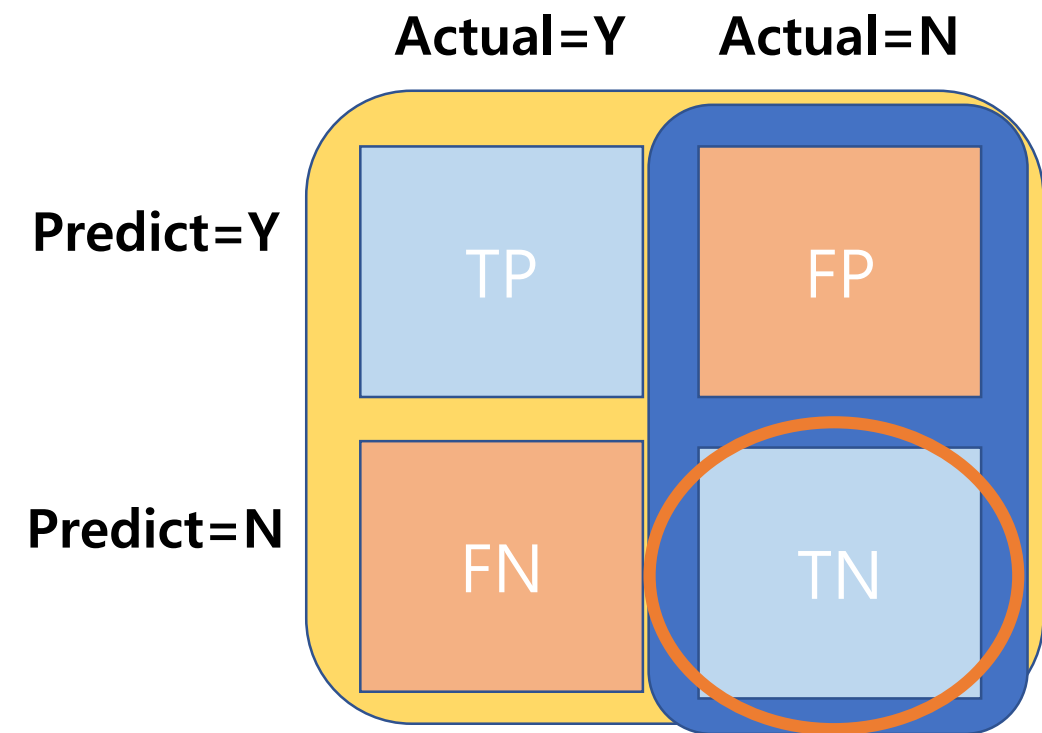
TP	FP
FN	TN

TP	FP
FN	TN

4-2 Metric

특이도(Specificity)

실제로 N(거짓)인 값을 N으로 예측한 확률



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

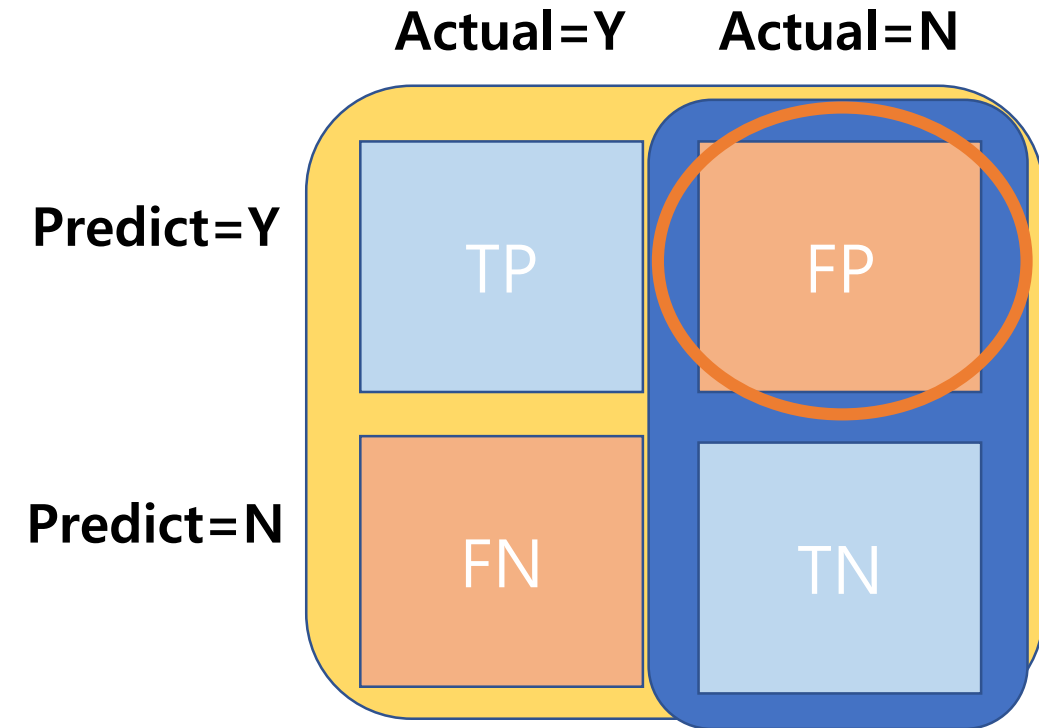
TP	FP
FN	TN

TP	FP
FN	TN

4-2 Metric

FPrate

실제로 No인데 예측이 Yes로 된 경우



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

TP	FP
FN	TN

TP	FP
FN	TN

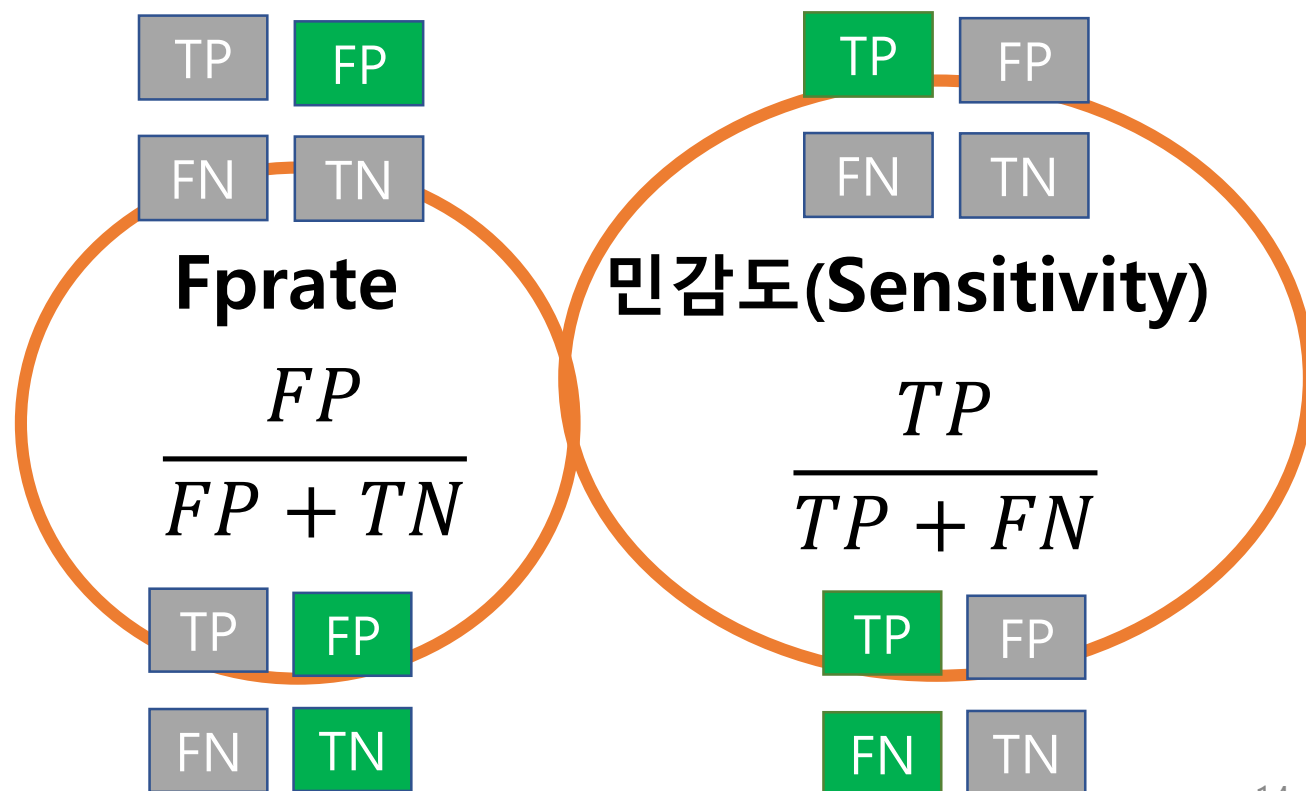
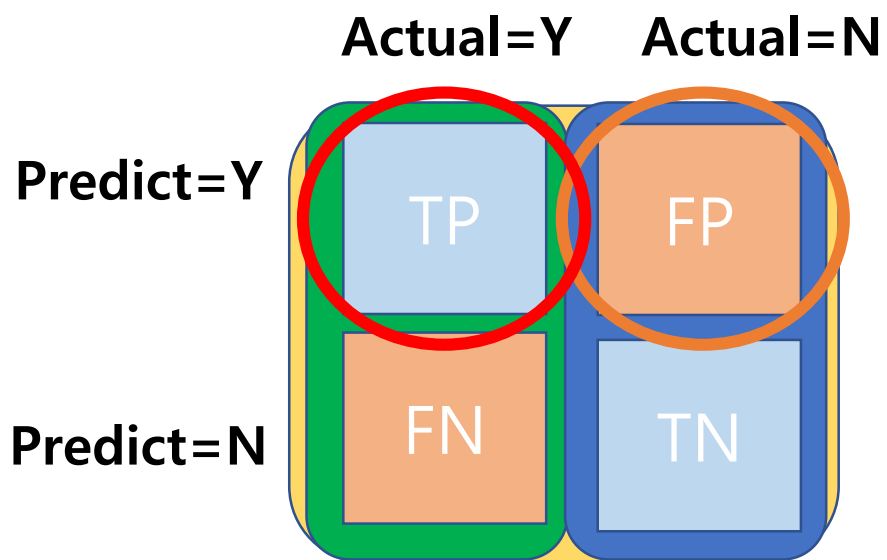
4-2 Metric

민감도(Sensitivity, Recall, TPrate)

실제로 Yes인데 예측이 Yes로 된 경우

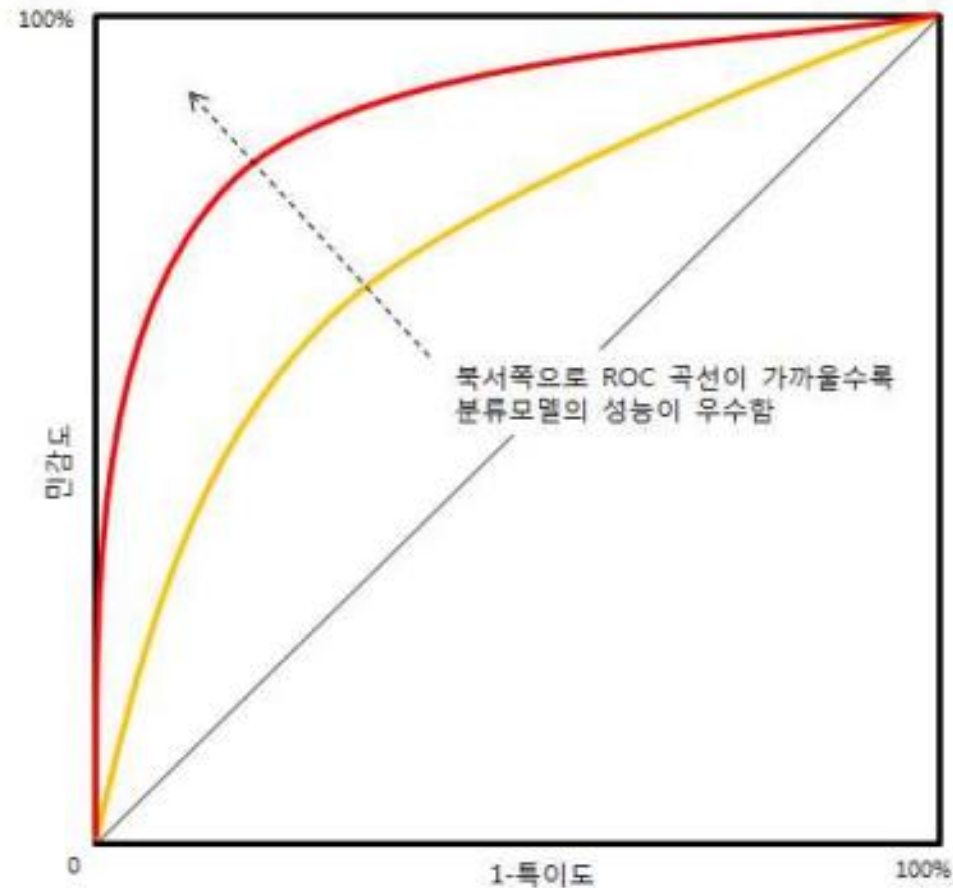
FPrate

실제로 No인데 예측이 Yes로 된 경우



4-2 Metric

ROC Curve



[그림 3-2] ROC 곡선

1-특이도는 Fprate이다.

