

평가 매트릭스

Classification Report(분류평가)

분류 Score

종류

1. 'accuracy_score'
2. 'f1_score', 'precision_score', 'recall_score' --> OvO vs. OvR
3. 'roc_auc_score' --> Macro-Average vs. Micro-Average
4. 'neg_log_loss'

OvO vs. OvR

- 1. OvO(One vs. One) : 이진분류(binary)
- 2. OvR(One vs. the Rest) : 다중분류(multi)

Macro-Average vs. Micro-Average

- 1. Macro-Average : 산술 평균
- 2. Micro-Average : 기하평균, 가중평균

Confusion Matrix(혼동 행렬)

혼동행렬 작성 방법

		Predicted		
		Negative (0)	Positive (1)	
Actual	Negative (0)	True Negative TN	False Positive FP (Type I error)	Specificity $= \frac{TN}{TN + FP}$
	Positive (1)	False Negative FN (Type II error)	True Positive TP	Recall, Sensitivity, True positive rate (TPR) $= \frac{TP}{TP + FN}$
Accuracy $= \frac{TP + TN}{TP + TN + FP + FN}$			Precision, Positive predictive value (PPV) $= \frac{TP}{TP + FP}$	F1-score $= 2 \times \frac{Recall \times Precision}{Recall + Precision}$

- 통계에서는 positive와 negative 위치 바뀜(주의하기!!!!)

평가지표

1. 재현율(Recall, 민감도(Sensitivity), TPR(True Positive Rate))

- 실제 p, 예측 p : 실제 p인 것 중에 예측값이 p인 것
- 예측모델은 대부분 재현율을 사용함
 - 양성(Positive)에 대해 틀리지 말아야 함(정상에 대해 관심 없음) --> FN중요
 - 실제와 예측이 다를 때 문제상황이 발생
- ex. 암환자의 예측, 범죄자 예측, 산불

$$\text{재현율} = \frac{TP}{FN + TP}$$

2. 정밀도(precision)

- 예측 p, 실제 p : 예측 p인 것 중에 실제값이 p인 것
- 양성(Positive)에 대해 틀리지 말아야 함(정상에 대해 관심 없음) --> FP중요
- ex. 스팸 메일 분류기

$$\text{정밀도} = \frac{TP}{FP + TP}$$

재현율과 정밀도의 관계

- Trade-off --> precision_recall_curve()로 확인 가능
- 서로 보완적인 관계 --> 둘 다 높아야 좋은 F1수치가 나옴

3. 정확도(accuracy) : 전체 중 모델이 올바르게 분류한(맞춘) 비율 -->분류모델은 많이 맞추는 것은 의미가 없음

4. 특이도(Specificity, TNR(True Negative Ratio))

- TNR(실제 N, 예측 N)

$$TNR = \frac{TN}{TN + FP}$$

- 예측모델은 TPR에만 관심이있기 때문에 전체 1의 비율을 기준으로 특이도를 빼서 FPR을 알고자 함
 - FPR=1-특이도

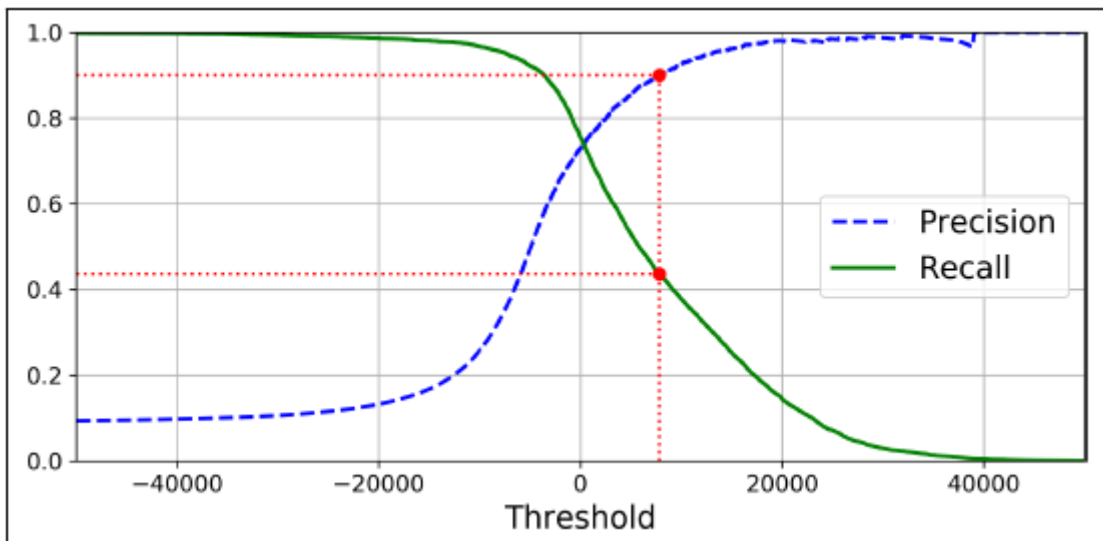
$$FPR = \frac{FP}{TN + FP}$$

임계치

- 임계치를 높일 수 록 양성학습(positive)는 감소한다
- 임계치를 낮추면 양성 학습을 많이 하게 됨 = positive를 더 잘 맞추게 됨

임계치와 재현율, 정밀도의 관계

- 임계치 감소
 - 재현율 증가, 정밀도 감소
 - 임계치를 낮추면 양성 학습을 많이 함 (p라고 답할 확률 증가)
 - TP 증가, TN 감소 --> FN 감소 --> 재현율 증가
- 임계치 증가
 - 재현율 감소, 정밀도 증가
 - 임계치 높이면 양성학습(Positive)이 줄어듦 --> FN증가 --> 재현율 감소(분모가 커짐)



적정 임계치 설정

- 재현율과 정밀도 모두 고려해야 함
 - 하나의 점수만 보게 되면, 임계치를 조작하여 극단적인 점수가 나오게 할 수 있기 때문 --> 모델의 신뢰성이 떨어짐
- 재현율, 정밀도 (Trade-off)
 - (임계치0 = 재현율100 ↔ 임계치1 = 정밀도100)
- 일반적으로 임계치는 0.5를 기준으로 ± 0.1 만 조정

F1 score

- 정밀도 재현률 조화
- 재현율과 정밀도의 점수가 모두 좋아야 F1_score ↑

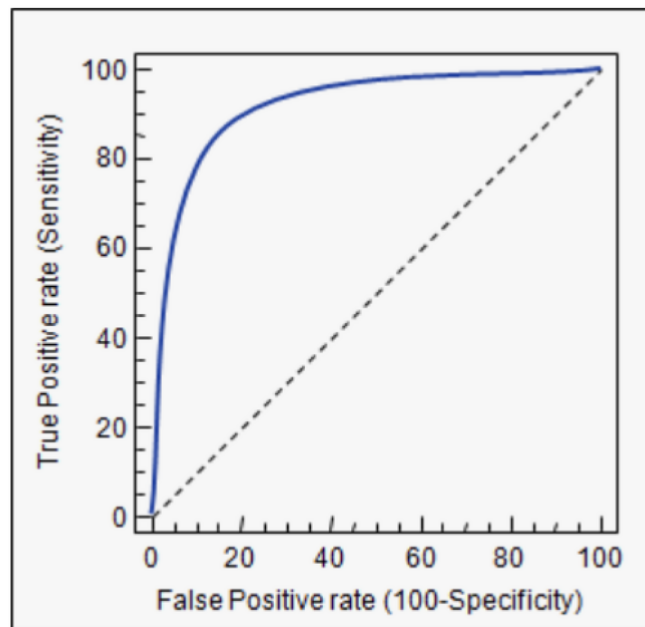
$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = 2 * \frac{precision * recall}{precision + recall}$$

ROC 곡선과 AUC score

ROC 곡선(수신자 판단 곡선)

개념

- Receiver Operating Characteristic Curve
- 민감도(Sensitivity)와 1-특이도(Specificity)로 그려지는 곡선을 의미
- 맞춘 것과 틀린 것에 대한 비율(양성 오답률)을 도식화
- FPR이 바뀔 때 TPR이 어떻게 바뀌는가 (X축 변화에 따른 Y축 변화) : TPR / FPR
- 임계치 0.5(직선) 밑으로 떨어진다는 것은 예측률이 낮다는 것
 - 임계치와 멀리떨어지는 것이 좋음: 직각에 가까운 형태가 가장 좋음(면적 1)



TPR(True Positive Ratio)

- Y축
- 재현율(Recall), 민감도(Sensitivity) --> 공식이 같은것, 동일한 것은 아님
- 양성비율 : p 전체를 대상으로, 예측모델이 정확히 p를 맞춘 비율()

FPR(False Positive Ratio)

- X축
- $FPR = 1 - \text{특이도(TNR)}$
 - 예측모델은 TPR에만 관심이있기 때문에, 전체 1의 비율을 기준으로 특이도를 빼서 FPR을 알고자 함

$$FPR = \frac{FP}{TN + FP}$$

- FPR이 0이 될 조건
 - $FPR = 0 \Rightarrow \text{임계치 } 1 \Rightarrow \text{TPR(재현율)} = 0$
- FPR이 1이 될 조건
 - $FPR = 1 \Rightarrow \text{임계치 } 0 \Rightarrow \text{TPR} = 1$
 - $TN = 0$

AUC score

개념

- ROC 곡선 밑의 면적
- 0~1의 값을 가짐
- ROC 곡선의 형태가 직각에 가까워져야 면적이 최대화(1)됨

