

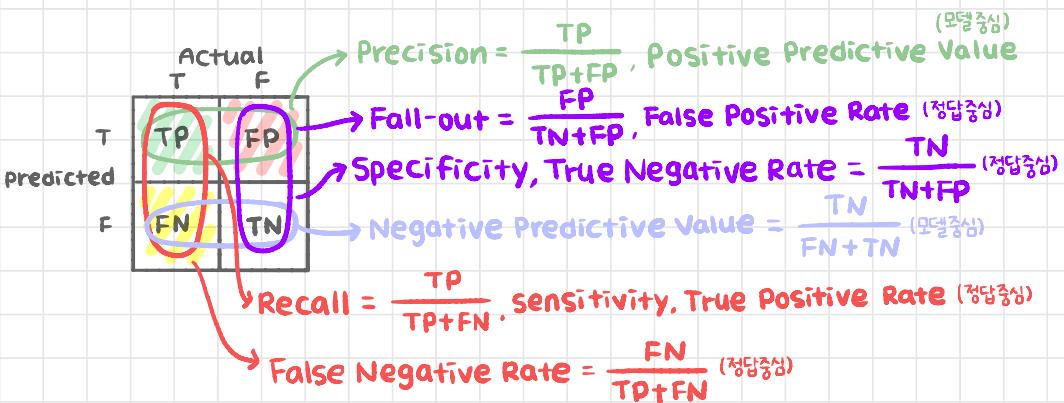
Evaluation for Classification

19.11.07 Sunmi ❤

* Confusion matrix

		실제정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}}$$



Precision (정밀도), PPV (Positive Predictive Value)

- 모델이 True라고 분류한 것 중에, 실제 True인 것의 비율

Recall (재현율), Sensitivity, hit rate

- 실제 True 중에 모델이 True로 분류한 비율

"Precision만 신경을 쓰면 모델이 인식해지고, Recall만 신경쓰면 모델이 헤퍼진다"

"Precision과 Recall은 상호보완적으로 사용할 수 있으며, 두 지표가 모두 높을수록 좋은 모델"

Accuracy

- TP, TN을 모두 고려하는 지표
- Label 불균형이 심할 때에 사용 주의해야 함

F1 Score

→ 산술 평균을 이용하는 것보다, 큰 비중이 끼치는 bias가 들어온다.

- Precision과 Recall의 조화평균
- Label 불균형이 심할 때 모델의 성능을 정확하게 평가 가능

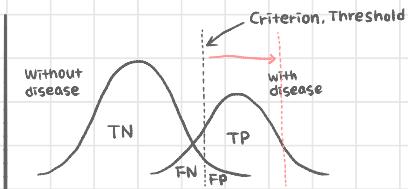
"Label 불균형이 심할 때 Accuracy는 지표로서 신뢰성을 잃는다."

Fall-out, FPR (False Positive Rate)

- 실제 False인 data 중에서 모델이 True라고 예측한 비율

$$\text{Fall-out (FPR)} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

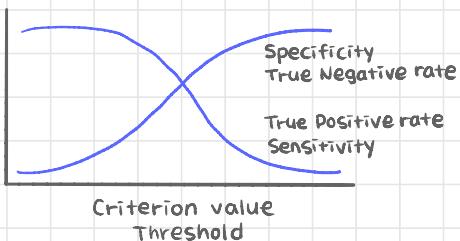
* Confusion Matrix with Histogram



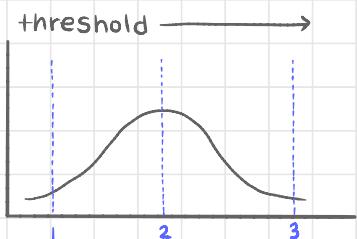
Threshold가 극단적으로 우측으로 이동하면?

- Precision ↑
- Recall (Sensitivity) ↓
- Specificity ↑
- Fall-out (False Positive Rate) ↓
- True positive Rate ↑
- True Negative Rate ↓

Specificity 와 Sensitivity 의 관계



두 distribution이 완벽히 일치한다고 가정



$$1. \text{ recall} = 100$$

$$\text{specificity} = 0$$

$$\text{fall-out rate} = 100$$

$$2. \text{ recall} = 50$$

$$\text{specificity} = 50$$

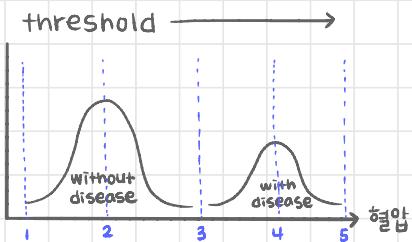
$$\text{fall-out rate} = 50$$

$$3. \text{ recall} = 0$$

$$\text{specificity} = 100$$

$$\text{fall-out rate} = 0$$

두 distribution이 겹치지 않는다고 가정



- recall = 100 (4축)
specificity = 0
fall-out rate = 100 (7축)
- True positive rate = 1
(100, 100)

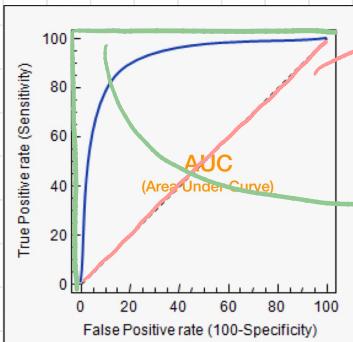
- recall = 100
specificity = 50
fall-out rate = 50

- recall = 100
specificity = 100
fall-out rate = 0

- recall = 50
specificity = 100
fall-out rate = 0

- recall = 0
specificity = 100
fall-out rate = 0

* ROC (Receiver Operating Characteristic) Curve



Actual True와 Actual False distribution이 완벽하게 같을 때
(feature의 class 변별능력 없음)

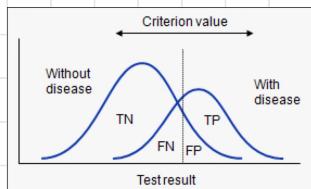
ROC curve는 45° 각도 직선
random하게 예측했을 때

Actual True와 Actual False distribution이 겹치는 영역 없이
완벽하게 분리될 때 ROC curve
(feature의 class 변별능력이 완벽)

ROC curve가 좁상단에 가까울수록 feature의 class 변별능력이 좋다.

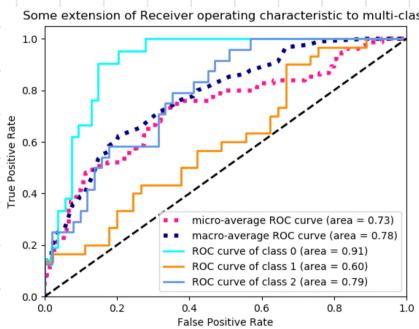
* ROC curve with Machine Learning

classifier를 만든다는 건, 두 개의 histogram을 그리고 Threshold를 정의하는 것



Histogram을 그렸다는 건 ROC curve를 그릴 수 있다는 것

→ 여러 ROC curve 비교를 통해 좋은 성능의 모델을 찾아낼 수 있다는 것!



AUC가 크다

- 모델이 계산한 probability를 바탕으로 그린 histogram들이 잘 분리되어 있다.
- 모델이 Threshold (Decision Boundary)에 덜 민감하다
- 안정적인 예측을 한다.

모델 선택에 ROC curve를 활용한다

- Decision Boundary에 상관없이 더 좋은 모델을 찾는다
- ganzi가 난다 😎