

기초 이론부터 실무 실습까지  
머신 러닝 익히기

# Part 10. 성능 평가

정 정 민

## Chapter 24. 성능 평가 metric

---

1. 성능 평가
2. 분류 문제 성능 추가 metric
3. 회귀 문제 성능 추가 metric

# 성능 평가

# 성능 평가는

---

- 머신 러닝 모델의 성능을 객관적으로 측정하고 비교하는 과정에서 사용
- 이를 통해 모델의 강점과 약점을 파악
- 성능 평가에 사용되는 지표를 **metric**이라고 함
- Metric에는 다양한 종류가 있으며
- 각각의 지표가 제공하는 정보가 다름
  - 목적에 맞는 올바른 **metric**을 선택해야 함
  - 또는 해석이 잘 되는 **metric**!
- 풀고자 하는 문제에 딱 맞는 **metric**이 존재하지만
- 적지 않은 경우로 **적절한 metric이 없을 수 있음**
  - 예를 들어, 고객이 만족하는 텍스트 prompt 기반 이미지 생성 모델의 성능은???
  - 특히 비지도 학습의 경우 문제의 목적에 딱 fit한 **metric**을 찾기 어려움
- 원초적이지만 시각화도 좋은 방법!

# 지금까지 살펴본 Metric은

---

## 지도 학습

- 분류 문제
  - 정확도 (Accuracy)
  - 혼동 행렬 (Confusion matrix)
  - 정밀도 (Precision)
  - 재현율 (Recall)
  - F1 score
- 회귀 문제
  - MSE (Mean Squared Error)

## 비지도 학습

- 군집화
  - SSE (Sum of Squared Error)
  - 실루엣 계수 (Silhouette Coefficient)
- 이상치 탐지
  - 분류 문제의 metric을 대안으로 활용

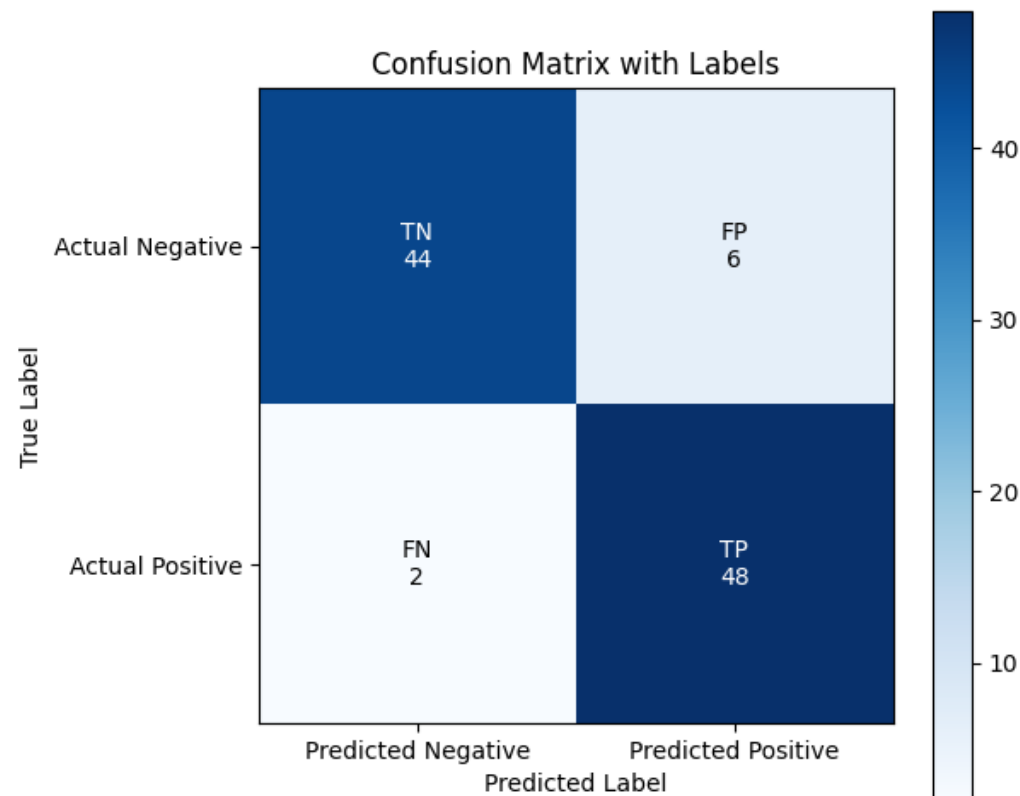
# 분류 문제 성능 추가 Metric

# Confusion Matrix

- 분류 문제에서 모델의 성능을 이해하고 해석하기 위한 중요한 도구
- 실제 레이블과 예측 레이블을 비교해 모델의 성능을 시각적으로 표현한 행렬

- 다음과 같은 4가지 요소로 구성
  - 진짜 양성 (True Positive, TP)
  - 거짓 양성 (False Negative, FN)
  - 진짜 음성 (True Negative, TN)
  - 거짓 음성 (False Negative, FN)
  - 이는 암기의 요소가 아니라 이름을 갖게 된 과정을 이해할 것!

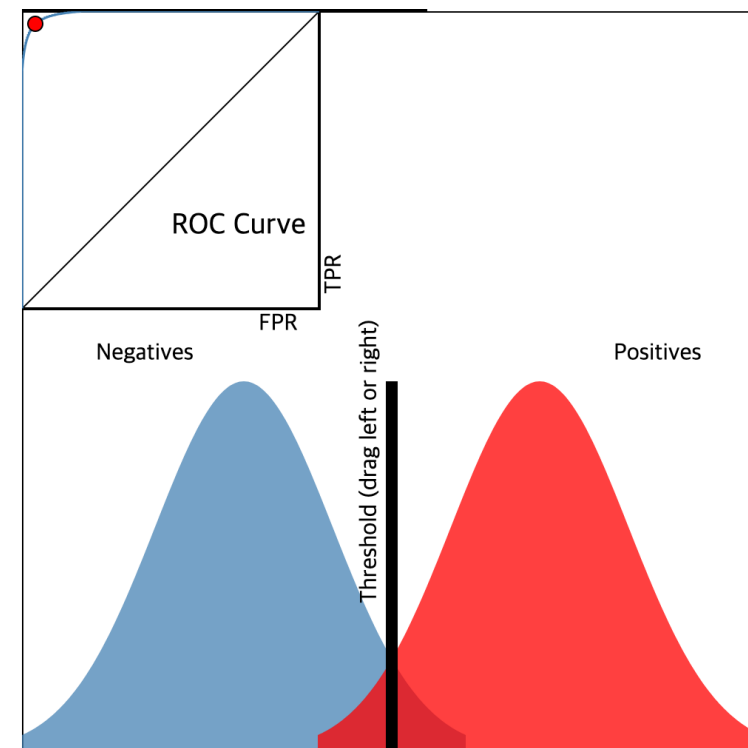
- $Accuracy = \frac{TN+TP}{TN+FP+FN+TP}$
- $Precision = \frac{TP}{FP+TP}$
- $Recall = \frac{TP}{FN+TP}$





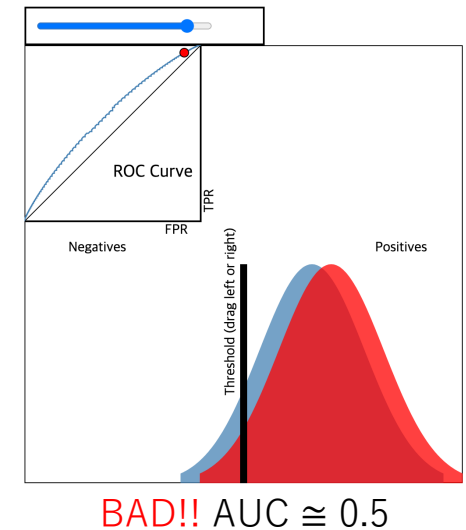
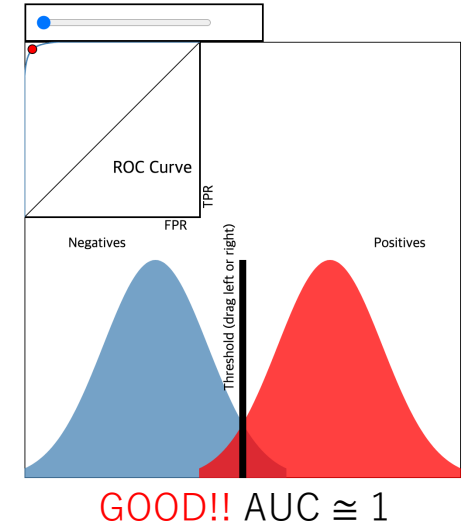
# ROC(Receiver Operating Characteristic) curve

- 이진 분류 문제에서 널리 사용되는 모델 성능 측정 도구
- 양성(positive)과 음성(negative)을 나누는 임계값의 변화에 따른 성능을 시각화 한 그래프
- 성능은 아래의 두 파라미터를 사용
  - x 축 : 거짓 양성률 (False Positive Rate, FPR)  $FPR = \frac{FP}{TN+FP}$ 
    - 전체 음성 중에서 실제로는 음성인데 양성이라고 잘못 찾은 비율
  - y 축 : 진짜 양성률 (True Positive Rate, TPR)  $TPR = \frac{TP}{FN+TP} = Recall$
- 왼쪽 위에 포인트가 존재하게 하는 임계치를 선택하는게 좋음!
- 즉, ROC 커브란
  - 머신러닝 모델이 양성을 최대한 많이, 잘 찾아내면서
  - 잡음에 의한 거짓 탐지를 최소화 하는지를 바탕으로
  - 적절한 임계치를 찾는 도구



# AUC(Area Under the Curve) 점수

- 음성과 양성을 잘 구분할수록 좋은 분류기!
- 그것이 ROC 커브 안에서도 확인이 가능
  - 음성과 양성을 잘 나눴다면 ROC 커브는 왼쪽 위 방향으로 그래프가 치우침
  - 그렇지 않다면  $y = x$  그래프에 수렴
- 즉, 좋은 분류기인지 아닌지에 따라
- ROC 커브의 모양이 바뀌고
- 그것을 수치로 나타내려면 ROC 커브의 아래 면적으로 표시할 수 있음
  - 그래서 이름이 Area Under the Curve!
  - 최대 : 1 최소 : 0.5
- 정리하면, AUC는 분류기의 전반적인 성능을 알 수 있는 척도로
- ROC 커브의 아랫 면적을 의미함



# 회귀 문제 성능 추가 Metric

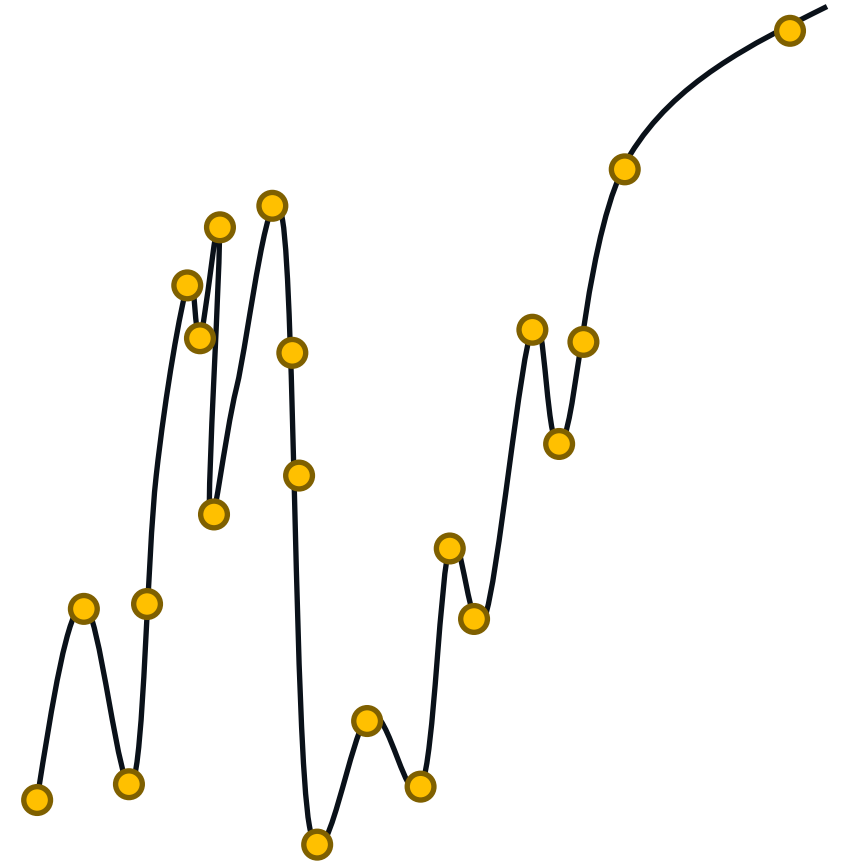
- 회귀 모델의 성능을 평가하는 통계적 지표
- 결정 계수(Coefficient of Determination)라고도 함
- 모델이 데이터의 변동성을 얼마나 잘 설명하는지를 나타냄

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

- $SS_{res}$  : 모델 예측 값과 실제 값 사이의 제곱 합
- $SS_{tot}$  : 실제 값들의 평균과 실제 값 사이의 차이 제곱 합
  - 이는 실제 데이터 내의 총 변동량을 나타냄
  - 모델이 설명해야 하는 '전체' 변동성의 양을 의미함
- $R^2 = 1$  : 모델이 데이터의 변동성을 완벽하게 설명함 > 모든 데이터 포인트가 회귀선에 정확히 놓여 있음
- $R^2 = 0$  : 변동성을 설명하지 못함 > 단순히 모델의 평균값을 예측
- $0 < R^2 < 1$  : 일부 변동성을 설명,  $R^2$ 이 클수록 설명력이 높다고 볼 수 있음

## $R^2$ 와 과적합

- $R^2$  값은 클 수록 설명력이 크다고 할 수 있음
- 하지만 **과적합의 위험성**을 주의해야 함
  - 더 많은 변수를 추가하면 모델이 비대해지고  $R^2$ 값은 자동으로 증가
  - 이때 과적합의 상태로 빠지는 것을 주의



$R^2 = 1$  그리고 과적합!!

**E.O.D**