

3-2-5 모형평가

모형평가

가장 적합한 모형을 선택하기 위해서 모형 평가 기준 필요

모형 평가기준

- 일반화 가능성
 - 같은 모집단 내의 다른 데이터에 적용하는 경우에도 안정적인 결과를 제공하는 것으로 의미
 - 데이터를 확장하여 적용할 수 있는지에 대한 평가 기준
- 효율성
 - 효율성은 분류 분석 모형이 얼마나 효율적으로 구축되었는지 평가
- 예측과 분류의 정확성
 - 효율적인 내용을 구축하였다 하더라도 실제 문제점이었을 때 정확하지 못한 결과물을 양산한다면 그 모형의 의미를 가질 수 없음

데이터 분할

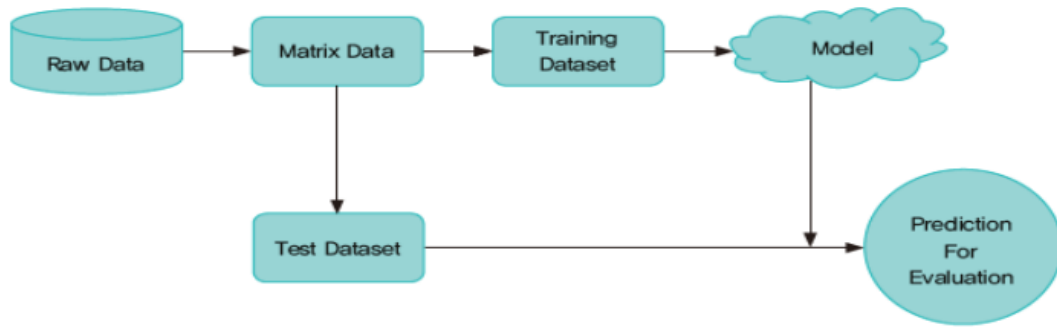
과적합을 방지하기 위해서 사용

- 과적합이란 : 주어진 데이터에서만 높은 성과를 보이고 다른 검증 데이터에는 잘 맞지 않는 현상
- 과소적합(Under Fitting) 모델이 너무 간단하여 정확도가 낮은 모델
- Train Data: 분석모형을 만들기 위한 학습용 데이터
- Validation Data: 학습된 데이터를 가지고 검증을 하는 데이터
- Test Data: 얼마나 잘 작동하는지 확인하기 위한 결과용 데이터

데이터 분할 방법

홀드아웃(hold-out)

- Raw data를 두 분류 분리하고 테스트 실시
 - 70%는 훈련용, 30%는 테스트용 자료로 사용
- Test Data는 검증용으로만 Model 할때는 Training Data만 사용



홀드아웃 방법

출처 <https://acdongpgm.tistory.com/88>

교차검증(Cross Validation)

주어진 데이터를 가지고 반복적으로 성과를 측정하고 그 결과를 평균한 것

Iteration / 데이터	A	B	C	D	E
1	Test	Train	Train	Train	Train
2	Train	Test	Train	Train	Train
3	Train	Train	Test	Train	Train
4	Train	Train	Train	Test	Train
5	Train	Train	Train	Train	Test

※ 5번 반복측정하고 각각의 반복측정 결과 평균

일반적으로 10-fold 교차검증 사용(k-fold)

부스트랩(Bootstrap)

평가를 반복해서 교차검증과 유사하나 훈련용 자료를 반복 재선정

- 관측치를 한 번 이상 훈련용 자료로 사용하는 복원추출법
- 비교적 작은 데이터 세트에 적합

데이터 분할 방식 비교

항목	홀드아웃 (Hold-out)	교차검증 (Cross Validation)	부스트랩 (Bootstrap)
데이터 분할 방식	학습/검증 데이터 한 번만 분할	데이터를 여러 개 fold로 나누어 반복 평가	복원추출로 여러 훈련 세트를 생성하여 반복 학습
샘플링 방법	무작위 분할 (비복원)	무작위 분할 (비복원, 다회전)	복원추출 (중복 허용)

항목	홀드아웃 (Hold-out)	교차검증 (Cross Validation)	부스트랩 (Bootstrap)
반복 횟수	1회	일반적으로 k회 (예: 5-fold, 10-fold)	수백~수천 회까지 가능
계산 비용	낮음	중간 (k배 반복)	높음 (많은 반복과 훈련 필요)
장점	빠르고 간단함	일반화 성능 추정에 신뢰도 높음	적은 데이터에서도 추정 가능 신뢰구간 계산 가능
단점	데이터 쏠림에 민감	계산량 증가, 데이터 편중 시 fold 품질 저하 가능	중복 데이터로 인한 과적합 가능, 계산 비용 큼
적용 사례	빠른 실험, 대규모 데이터	일반적인 모델 성능 검증	통계적 신뢰성 추정, 작은 샘플에서의 불확실성 분석에 적합

분류 모형 평가를 위해 범주형 변수의 대표적으로 사용되는 방법

- 오분류표
- ROC(Receiver Operating Characteristic) 그래프
- 이익 도표(Gain chart)
- 향상도 곡선(Lift Curve) 등

오분류표(혼동 행렬, confusion matrix)

		예측치	예측치	
		True	False	합계
실제값	True	TP	FN	P
실제값	False	FP	TN	N
	합계	P'	N'	P+N

정리

Confusion matrix		예측값		
		TRUE	FALSE	
실제값	TRUE	40 (TP)	60 (FN) Type II Error	Sensitivity TP / (TP+FN)
	FALSE	60 (FP) Type I Error	40 (TN)	Specificity TN / (TN+FP)
	Precision TP / (TP+FP)		Negative Predictive Value TN / (TN + FN)	Accuracy (TP+TN) / (TP+TN+FP+FN)

실Sen, 예Pre

T/F

P/N

실제 == 예측 : True

True 예측 : Positive

실제 != 예측 : False

False 예측 : Negative

T

P

T

N

F

P

F

N

confusion matrix		실제값	
		Y	N
예측값	Y	True Positive	False Positive
	N	False Negative	True Negative

- 정분류율, 정확도(Accuracy, AR) = $(TN + TP) / (TN + TP + FN + FP)$

Accuracy

		Predicted	
		Spam	Not
Actual	Spam	600 (TP)	300 (FN)
	Not	100 (FP)	9000 (TN)

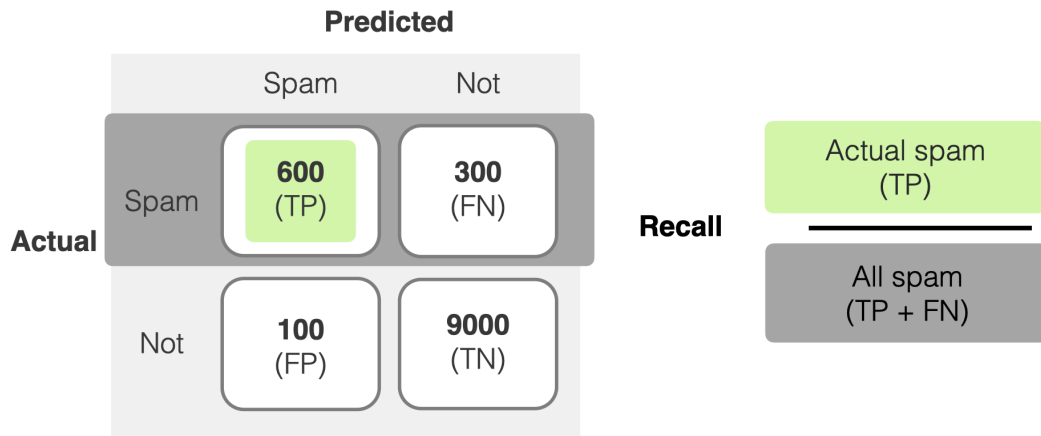
Accuracy = $\frac{\text{True predictions (TP + TN)}}{\text{All predictions (TP + TN + FP + FN)}}$

출처 <https://www.evidentlyai.com/classification-metrics/confusion-matrix>

- 오분류율(Error Rate) = $1 - \text{Accuracy} = (FN + FP) / (TN + TP + FN + FP)$
- 민감도(Sensitivity, TPR) = $TP / (TP + FN)$ # True인 케이스에서 True로 예측한 비율
 - TPR(True Positive Rate)
 - 분류 모델이 실제 양성 클래스(positive class)를 얼마나 잘 찾아냈는지를 나타내는 지표
 - 다른 말로 재현율(Recall)이라고도 불러
 - 공식은 $TP / (TP + FN)$ 이고

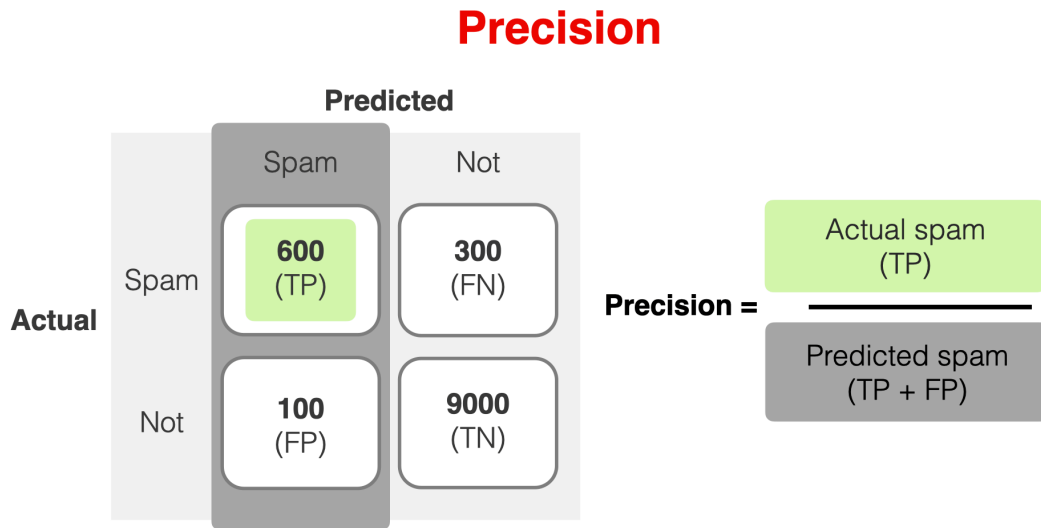
- 값이 높을수록 실제 양성(positive)을 잘 맞췄다는 의미
- 예: 질병 진단에서 TPR이 0.95라면, 실제 환자 100명 중 95명을 정확히 감지했다는 의미

Recall



-
- 특이도(Specificity, TNR) = $TN / (TN + FP)$
 - TNR(True Negative Rate): 실제 음성(Negative)인 샘플 중에서 모델이 올바르게 음성으로 예측한 비율
 - 다른 이름으로 특이도(Specificity)라고도 부름
 - 공식은 $TN / (TN + FP)$ 이고, False Positive를 얼마나 잘 줄였는지를 나타냄
 - 주로 의료 진단, 보안 탐지 등에서 TPR과 함께 사용
- FPR = 1 - Specificity # False인 케이스에서 True로 예측한 비율
 - FPR(False Positive Rate)는 실제로는 음성(Negative)인데 모델이 잘못 양성(Positive)으로 예측한 비율
 - 즉, 거짓 경보(false alarm)가 얼마나 자주 발생하는지 측정하는 지표
 - 공식은 $FP / (FP + TN)$
 - 이는 TNR(특이도)의 보완값, $FPR = 1 - TNR$
 - 질병 스크리닝: 병이 없는 사람을 환자라고 하면 불필요한 공포와 검사 비용 발생
- 정밀도(Precision) = $TP / (TP + FP)$
 - Precision(정밀도)는 모델이 양성(Positive)이라고 예측한 것들 중에서 실제로도 양성인 비율
 - 즉, 예측한 양성 중에 진짜가 얼마나 있냐를 보는 지표
 - 공식은 $TP / (TP + FP)$ 이고,
 - 거짓 양성(False Positive)을 얼마나 줄였는지가 핵심

- 특히 "양성 예측의 신뢰도"를 평가할 때 사용



- **F1 Score (F_1)** = $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$
 - **F1 Score**는 Precision(정밀도)과 Recall(재현율)의 조화 평균(harmonic mean)으로 계산되는 지표
 - 두 값을 **균형 있게 고려해서**,
 - 하나가 높고 하나가 낮을 경우 전체 성능을 냉정하게 낮게 평가
 - 공식은

$$F1 = 2 \times \frac{P \times R}{P + R}$$

- 불균형 데이터셋(binary imbalanced)에서 특히 유용하게 쓰임
 - Precision(정밀도)과 Recall(재현율)의 **균형을 측정하는 지표**
 - 정밀도와 재현율, 두 값이 **모두 높을수록** F1 Score도 높음
 - 둘 중 하나라도 낮으면 F1도 낮게 나옴 → **냉정한 평가자**

예시로 이해하기

값	예시 1	예시 2
Precision	0.9	0.5
Recall	0.9	0.5
F1 Score	0.9	0.5

- 하지만 Precision = 1.0, Recall = 0.1 이면?

$$F1 = 2 \times \frac{1.0 \times 0.1}{1.0 + 0.1} = 0.18$$

- → 한쪽만 높으면 안 통함. 둘 다 잘해야 F1이 높아짐!

F-β Score: F1의 일반형

요약

F-β 점수(F-beta score)는 Precision(정밀도)과 Recall(재현율) 사이의 중요도를 조절해서 조화 평균을 구하는 지표

- β(베타) 값을 통해 Recall을 더 중요하게 볼지, Precision을 더 중요하게 볼지 결정
 - β > 1이면 Recall 중심
 - β < 1이면 Precision 중심
 - β = 1이면 F1 Score와 같음

F-β Score 정의

공식

$$F_{\beta} = (1 + \beta^2) \times \frac{Precision \times Recall}{\beta^2 \times Precision + Recall}$$

- **Precision** = TP / (TP + FP)
- **Recall** = TP / (TP + FN)
- **β (베타)**: Precision 대비 Recall의 중요도 계수

해석

- **β > 1** → Recall에 더 높은 가중치
(예: 질병 진단, 사기 탐지처럼 놓치면 안 되는 상황)
- **β < 1** → Precision에 더 높은 가중치
(예: 스팸 필터링, 거짓 양성 줄이기 중요한 경우)
- **β = 1** → F1 Score (Precision과 Recall을 동일하게 고려)

예시 비교

β 값	의미	중심 지표
0.5	정밀도(Precision) 우선	FP 줄이기
1.0	균형 평가 (F1 Score)	균형형
2.0	재현율(Recall) 우선	FN 줄이기

F- β 점수가 중요한 상황

- 클래스 불균형 데이터셋
- 양성 클래스 놓치면 치명적인 문제 ($\beta \uparrow$)
- 잘못된 양성 예측 줄이는 게 핵심인 문제 ($\beta \downarrow$)
- 모델 성능 비교에서 Precision과 Recall의 가중치를 조절 가능

관련 지표 비교

지표	역할	공식
Precision	예측한 양성 중 진짜 양성 비율	$TP / (TP + FP)$
Recall	실제 양성 중 예측에 성공한 비율	$TP / (TP + FN)$
F1 Score	Precision과 Recall의 균형 평가 지표	$2 \times (P \times R) / (P + R)$
Accuracy	전체 예측 중 맞춘 비율 (불균형에 약함)	$(TP + TN) / (TP + FP + FN + TN)$

TNR vs TPR

구분	의미	잘 맞추는 대상	공식
TPR	민감도 (sensitivity)	실제 양성 (positive)	$TP / (TP + FN)$
TNR	특이도 (specificity)	실제 음성 (negative)	$TN / (TN + FP)$

두 값은 항상 trade-off 관계가 있음 → ROC Curve에서 함께 사용됨

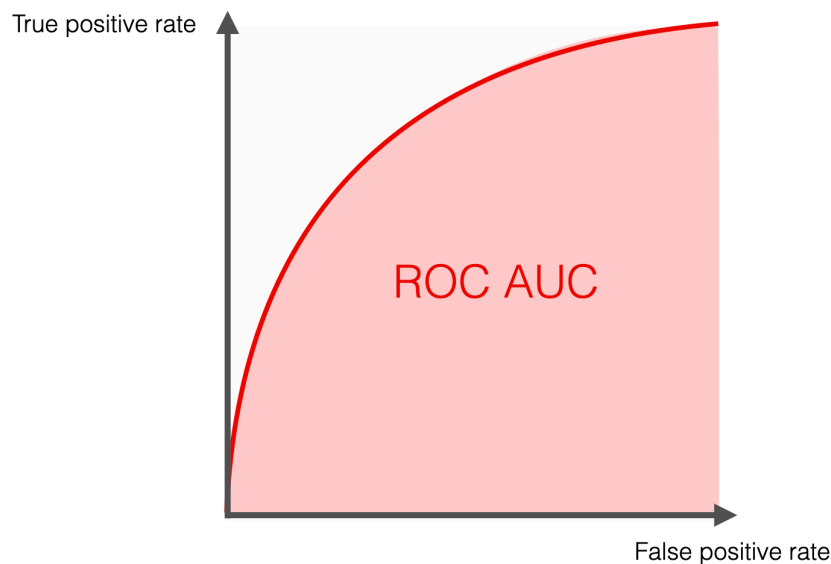
ROC 커브(Receiver Operating Characteristic Curve) 그래프

이진 분류에서 모형의 성능을 평가하기 위해 많이 사용되는 척도

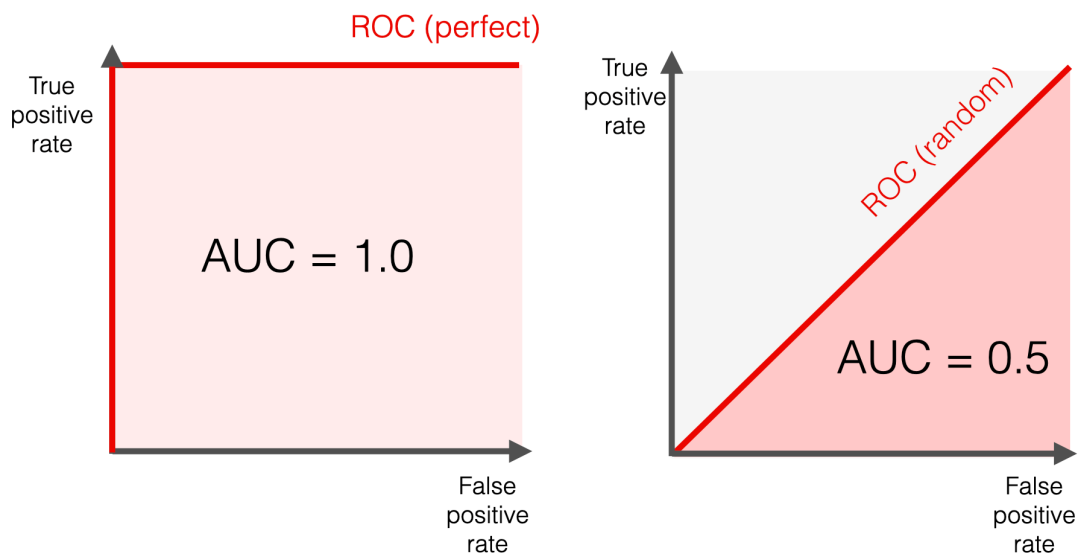
- 가로축을 FPR(False positive Rate, 1-특이도) 값으로 두고
- 세로축을 TPR(True Positive Rate, 민감도) 값으로 두어 시각화한 그래프
- 그래프가 왼쪽 상단에 가깝게 그려질수록 올바르게 예측한 비율은 높고, 잘못 예측한 비율이 낮음
- 임계값을 직접적으로 그래프에 표시하지는 않지만 커브 위의 각 점은 특정 임계값에서의 **FPR(위양성률)**, **TPR(민감도)** 조합
 - 즉, 여러 임계값을 변화시키며 점을 찍은 결과가 ROC 커브
 - "임계값(threshold)"은 분류 모델이 출력한 **예측 확률을 이진 분류(1 또는 0)로 바꿀 때 기준이 되는 값**
 - 예측 확률이 이 값보다 크면 1(양성), 작으면 0(음성)으로 분류
 - 기본값은 보통 **0.5**지만, 상황에 따라 조절하면 **민감도나 정밀도**를 개선할 수 있음
 - 임계값은 단순하지만, **모델의 판단 기준선을 바꾸는 레버**이기 때문에 매우 중요

따라서 **AUC**(Area Under ROC Curve, ROC 곡선 아래의 면적)가 값이 클수록 성능이 좋음

- **TPR(True Positive Rate, 민감도)**: 1인 케이스 대한 1로 예측한 비율
 - 모든 스팸 이메일 중 스팸으로 올바르게 분류된 이메일의 비율
- **FPR(False Positive Rate, 1-특이도)**: 0인 케이스에서 1로 잘못 예측한 비율
 - 정상적인 이메일이 스팸으로 잘못 분류된 비율



출처 <https://www.evidentlyai.com/classification-metrics/explain-roc-curve>

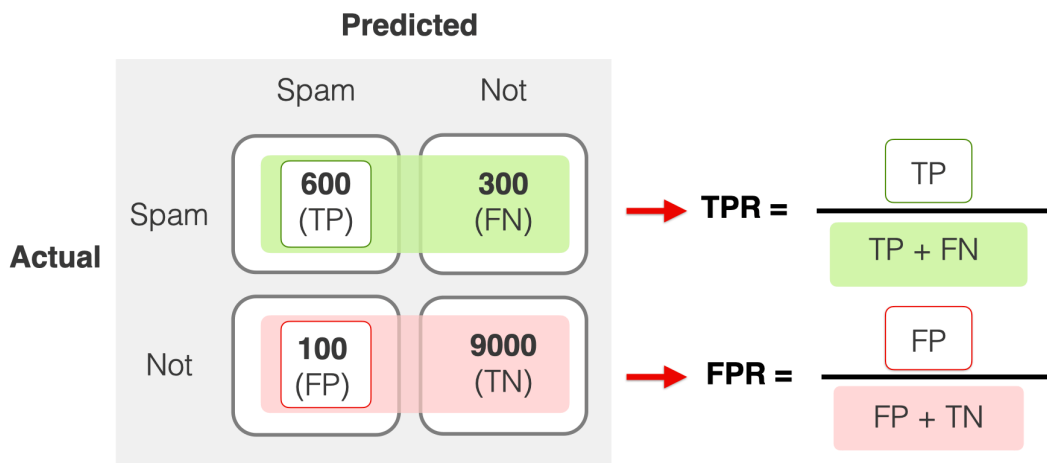


TPR과 FPR

ROC 곡선은 다양한 분류 임계값에서 참 양성률(TPR)과 거짓 양성률(FPR)을 비교한 그래프

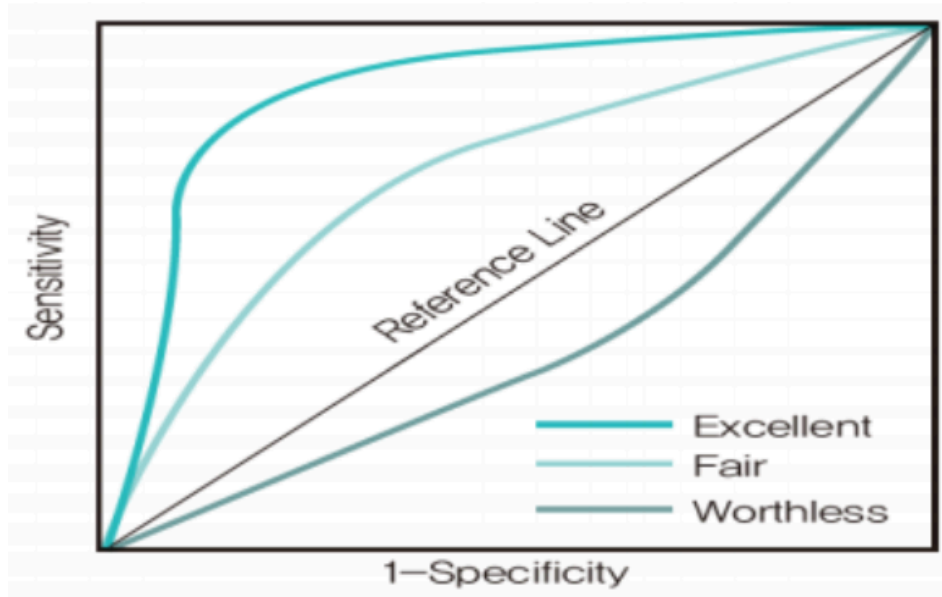
		Predicted	
		Spam	Non-spam
Actual	Spam	600 (True positive)	300 (False negative)
	Non-spam	100 (False positive)	9000 (True negative)





AUC를 이용한 정확도 판단 기준:

AUC 범위	평가 등급
0.9 ~ 1.0	Excellent
0.8 ~ 0.9	Good
0.7 ~ 0.8	Fair
0.6 ~ 0.7	Poor
0.5 ~ 0.6	Fail



가로축은 1-특이도 세로축은 민감도

요약:

- **ROC 그래프(Receiver Operating Characteristic Curve)**는 **이진 분류 모델의 성능을 시각화**하는 도구
 - X축은 **FPR(False Positive Rate)**, Y축은 **TPR(True Positive Rate / Recall)**로 구성
 - 모델의 **임계값(threshold)**을 조절하면서 TPR과 FPR이 어떻게 바뀌는지를 보여주는 곡선
 - 면적(AUC, Area Under Curve)이 클수록 좋은 모델을 의미

ROC 그래프 개념

1. 무엇을 그리는 그래프인가?

- 다양한 **임계값(threshold)**에 대해 모델의 성능(민감도 vs. 오탐률)을 시각화함
- 이진 분류 모델의 예측 확률을 기준으로, threshold를 바꿔가며 TPR과 FPR을 계산

2. 축 설명

축	의미
X축	FPR (False Positive Rate): $FP / (FP + TN)$
Y축	TPR (True Positive Rate): $TP / (TP + FN)$, 즉 Recall

ROC 곡선의 해석

- 좌측 상단에 가까울수록 좋은 모델 (TPR↑, FPR↓)
- 대각선은 완전 무작위 예측 모델 (AUC = 0.5)
- AUC가 1.0에 가까울수록 완벽한 분류기, 0.5에 가까울수록 의미 없음

ROC 곡선 예시

- 좋은 모델: 곡선이 좌측 상단에 치우쳐 있고 AUC \approx 0.9 이상
- 나쁜 모델: 곡선이 대각선에 가까움 (AUC \approx 0.5)
- 이상한 모델: AUC < 0.5 \rightarrow 예측 결과를 뒤집으면 오히려 좋아짐

ROC vs Precision-Recall 곡선

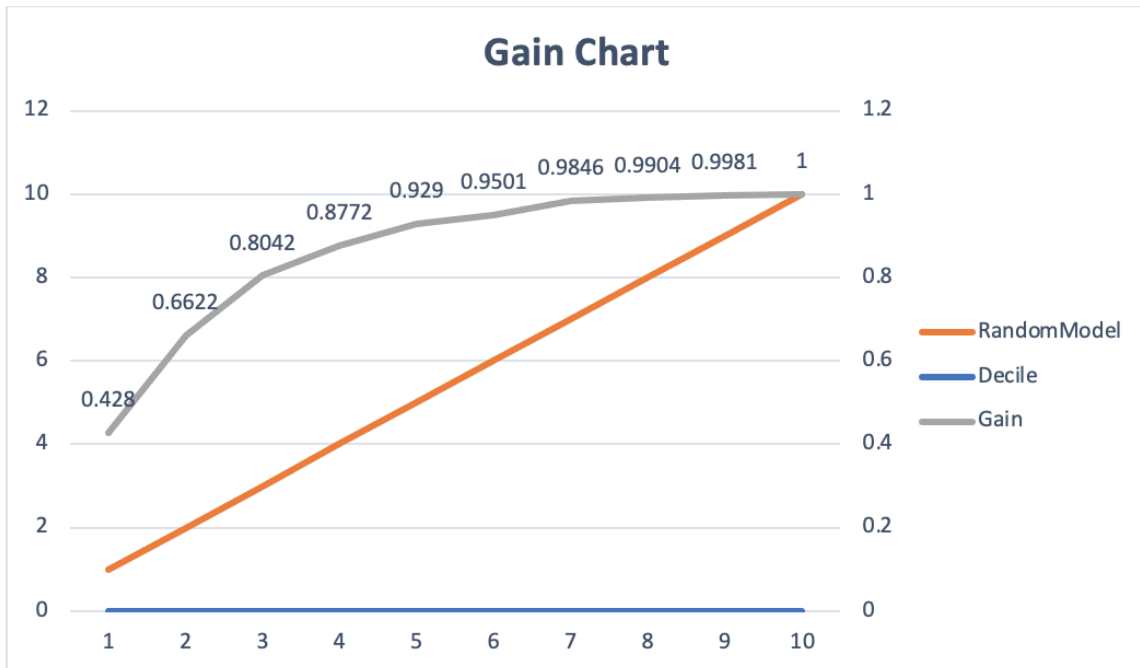
항목	ROC Curve	Precision-Recall Curve
X축	FPR	Recall
Y축	TPR	Precision
불균형 데이터에 강함?	보통 아님	Yes! Precision-Recall이 유리함

즉, 클래스가 불균형하면 PR curve가 더 실용적이야.

ROC는 전체 분포를 기반으로 해서, 음성 샘플이 많은 경우 FPR이 지나치게 낮게 나올 수 있어.

이익도표(gain chart)

- 분류된 관측치에 대해 얼마나 예측이 잘 이루어졌는지 나타내기 위해
 - 임의로 나눈 각 등급별로 반응 검출율, 반응률, 리프트 등의 정보를 산출하여 나타내는 도표
 - 이익 도표의 각 등급은 예측 확률에 따라 매겨진 순위이기 때문에 상위 등급에서는 더 높은 반응률을 보이는 것이 좋은 모형이라고 평가
 - 모델의 예측력이 높을수록 그래프가 대각선 위로 크게 벌어짐
 - 이게 모델의 "선택 능력"



출처 <https://www.geeksforgeeks.org/data-science/understanding-gain-chart-and-lift-chart/>

이익 도표 정리

게인 차트(Gain Chart)는 이진 분류 모델이 **양성(Positive) 표본을 얼마나 잘 선별해내는지** 시각적으로 보여주는 도구.

원래 마케팅, 금융 등에서 **제한된 자원으로 최대 효과를 내기 위해** 도입됐고, 지금은 머신러닝 모델 평가의 핵심 보조 지표로 사용

x축은 **전체 데이터 중 상위 몇 %를 선택했는가**, y축은 **그 안에 포함된 양성의 누적 비율**을 의미

모델의 예측력이 높을수록 그래프가 대각선 위로 크게 벌어짐. 이게 모델의 "선택 능력"

도입 배경: 왜 게인 차트를 쓸까?

1. 리소스는 한정돼 있고, 우리는 양성만 원한다

- 마케팅 예시: 전체 고객 10,000명 중 양성 반응 보일 사람이 500명뿐이라면?
- 모두에게 캠페인 돌리기엔 예산 부족!
- 그래서 모델을 써서 "누가 양성일 가능성이 높을까?" 점수를 매기고, **점수 높은 순으로 일부만 타겟팅**하는 전략을 씀.

2. 정확도는 이런 상황에서 무용지물

- 정확도는 양성/음성 클래스 비율이 불균형하면 쓸모 없음
- 예: 양성 비율 5%일 때, 모두 음성이라 해도 정확도 95%

- → 그래서 "순위 기반 평가" 지표가 필요하고, 그중 하나가 게인 차트임

게인 차트의 의미와 직관

1. 무엇을 보여주는가?

- x축 기준으로 상위 몇 %를 선택했을 때,
- 실제 양성(정답)을 몇 %나 커버할 수 있었는지를 보여줌
- 즉, "몇 명 골라야 진짜 고객을 몇 명 확보할 수 있을까?"의 답

2. 대각선은 무작위(Random) 모델

- 주황색 대각선: 점수를 랜덤하게 뽑았을 때의 기대 성과
- 회색 곡선(모델)이 이보다 위에 있으면 의미 있음
- 곡선이 더 빨리 위로 치솟을수록 모델이 "진짜 양성"을 더 잘 골라냄

3. 모델의 순위화 능력을 보여주는 것

- 단순히 예/아니오 맞췄냐가 아니라
- "누가 더 가능성이 높은지 잘 구분했는가?"에 초점을 둬

x축과 y축 설명

✓ x축: 누적 인원 비율 (예측 점수 기준 상위 %)

- 모델이 부여한 점수로 정렬한 후
- 상위 10%, 20%, ..., 100%까지 얼마나 많은 데이터를 포함했는가?
- 즉, "몇 %까지 골라봤을 때..."라는 관점

예:

- $x=0.3$ 이면 → 점수 상위 30% 표본을 대상으로 삼았다는 뜻

✓ y축: 누적 정답(양성) 비율

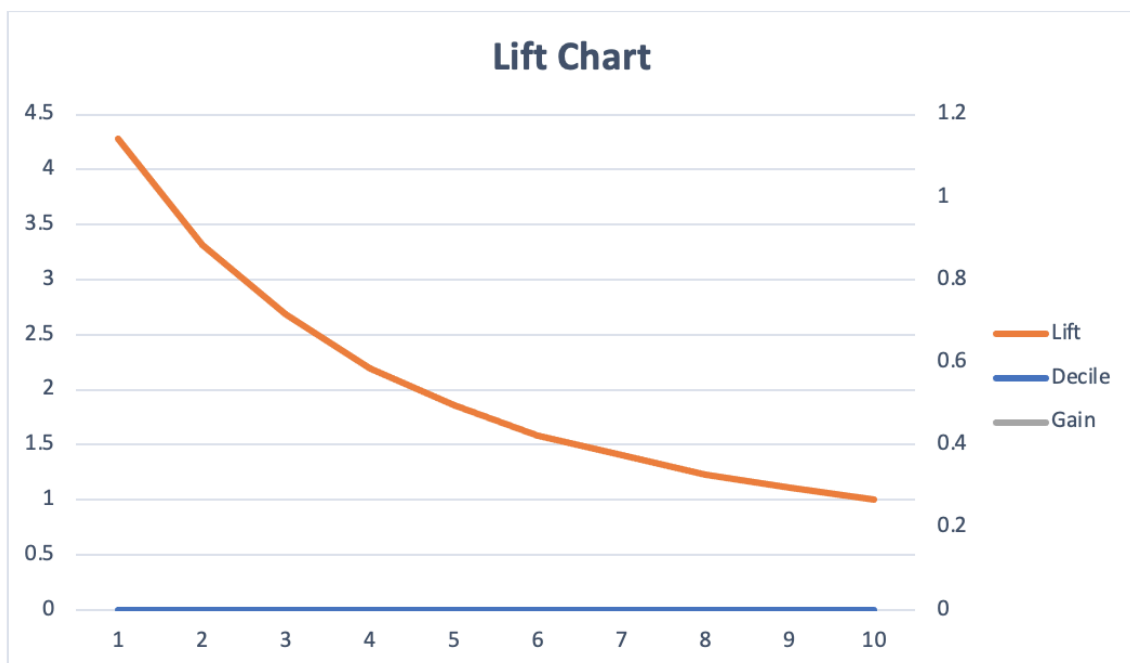
- 해당 x% 안에 실제 양성 표본이 몇 %나 들어있는지 누적 계산
- 예:
 - $x=0.3$ 일 때 $y=0.7$ → 상위 30% 표본 안에 전체 양성의 70%가 포함되어 있다는 뜻
 - 이건 매우 좋은 성과야!

실제 해석 예시

x (상위 %)	y (양성 누적 비율)	의미
10%	42.8%	상위 10%만 타겟팅해도 전체 양성의 절반 가까이 잡음
30%	80.4%	상위 30%만 골라도 양성의 80%를 확보
50%	92.9%	상위 절반으로 거의 전부 확보됨
100%	100%	전체 선택하면 당연히 양성도 전부 포함됨 (당연)

향상도 곡선(Lift Curve)

- 모델이 무작위(Random) 선정 대비 얼마나 효율적으로 목표를 찾아내는지(비율·배수)를 각 구간(데실 등) 별로 나타낸 그래프



- 1데실(상위 10%)은 모델이 가장 확신하는 집단 → 양성 비율이 매우 높아 무작위 대비 배수가 크다.
- 뒤로 갈수록(2~10데실) 모델 점수가 낮아져 양성 비율이 떨어짐 → 무작위 비율(기본값 1)에 가까워진다.
- 10데실(하위 10%)은 거의 무작위와 동일하므로 Lift ≈ 1이 된다.

그래서 좋은 모델일수록 첫 데실에서 Lift 값이 크게 튀고, 뒤로 갈수록 완만하게 1에 수렴하는 형태를 보임

리프트 곡선 정리

리프트 곡선(Lift Curve)은 모델이 무작위(Random) 선정 대비 얼마나 더 효율적으로 양성을 골라내는지를 배수(Lift)로 시각화한 그래프야.

이는 자원이 제한된 마케팅·의료·사기탐지 등에서 효율적인 타겟팅 의사결정을 위해 도입됐고, 모델의 순위화 성능을 평가하는 중요한 도구로 사용돼.

x축은 상위 점수 기준으로 얼마나 많은 데이터를 포함했는가(% 기준),

y축은 그 구간의 Lift 값 = 양성 비율 / 무작위 비율을 의미해.

Lift Curve의 도입 배경

1. 자원은 한정, 타겟은 희소 → 똑똑하게 골라야 함

- 실무에서 양성(Positive class)이 전체의 5~10% 이하인 경우가 많음 (예: 구매 고객, 사기 거래 등)
- 모든 사람에게 마케팅을 걸거나 검사를 하긴 어렵고,
- “누가 가능성 있는지” 점수를 매겨 상위만 타겟팅하는 전략이 필요함

2. 모델이 무작위보다 얼마나 더 잘 고르나?

- 단순히 “몇 명 맞췄나”보다,
- “그냥 랜덤으로 골랐을 때보다 얼마나 낫나”가 중요해짐
- → 이 배수 효과(Lift)를 시각화한 것이 바로 **Lift Curve**

Lift Curve의 의미

1. Lift = 모델이 무작위보다 몇 배 잘 골랐는가?

$\text{Lift}(k) = \frac{\text{양성 비율 @ } k}{\text{전체 양성 비율}}$

$\text{Lift}(k) = \frac{\text{전체 양성 비율}}{\text{양성 비율 @ } k}$

예시:

- 전체 양성 비율 = 10%
- 상위 10%에서 양성 비율 = 40% → $\text{Lift} = 40\% / 10\% = 4.0$
- 즉, 모델 점수 기준 상위 10%만 골라도 무작위보다 4배나 많이 양성을 포함함

2. 리프트 곡선은 데실별 성과를 보여줌

- 각 구간(10%, 20%, ...)별 Lift 값을 계산하여 그래프로 표현
- 보통 초반 데실(1~3데실)은 Lift가 크고, 뒤로 갈수록 1에 수렴

x축과 y축 설명

✓ x축: 누적 비율 (%) – 모델 점수 기준으로 상위 몇 %까지 포함했는가

- 예: $x=0.3$ → 모델이 확신하는 상위 30% 데이터까지 포함함
- 모델의 예측 확률 점수 기준으로 내림차순 정렬했을 때의 누적 구간

✓ y축: 리프트(Lift) 값

- 해당 누적 구간에서의 **양성 비율 / 전체 양성 비율**
- $y=4.0 \rightarrow$ 해당 구간에서 양성 비율이 전체 평균보다 4배 높다는 의미
- $y=1.0 \rightarrow$ 모델이 무작위 수준이라는 의미 (선택 능력 없음)

Lift Curve로 얻는 인사이트

상위 % (x)	Lift 값 (y)	해석
10%	4.28	무작위보다 4.28배 효율적인 타겟팅
20%	3.3	상위 20%까지는 여전히 3배 이상의 성과
50%	1.5	상위 절반까지는 성능 유지, 이후는 거의 무작위
100%	1.0	전체 다 보면 무작위와 같아짐 (당연)

실무 활용

- **캠페인 대상 최적화:** Lift>2인 구간까지만 타겟팅
- **모델 비교 지표:** 곡선이 더 높고 오래 유지되면 우수
- **ROI 최적화:** 효율이 높은 데실에 자원 우선 배분

정리하자면

- **Lift Curve**는 무작위 대비 **모델의 상대적 효율**을 배수로 시각화한 그래프
- x축은 상위 몇 %까지 포함했는지, y축은 그 구간의 Lift
- $y=1$ 을 기준으로 위에 있으면 의미 있는 모델, 아래면 "쓰지 마라" 수준
- 초반 데실에서 높게 튀는 것이 좋은 모델의 특징

종료

