

# Precision & Recall

## + 직관적인 이해

- **accuracy의 문제점**

- 실제 양성(암)과 음성(정상)의 분포의 차이가 극심할 때, *Accuracy*는 부적합함. (예, 암=1%, 정상=99%)
- 실제로는 암환자를 예측하는게 중요함. (*TP*나 *FN*을 찾는게 중요)
- 하지만, *Accuracy*는  $TP + TN / ALL$ 에서 *TN*의 영향이 너무 큼

- **precision = ‘예측이 양성’인 것 중에서 ‘실제 양성’인 비율**

- 내가 예측한 것 모두 암이어야지 좋은 결과 -> (보수적인 예측)일수록 좋음

- **recall = ‘실제 양성’인 것 중에 ‘정확하게 예측’한 비율**

- 실제 암환자를 놓쳐서는 안됨 -> (후한 예측)일수록 좋음

- **Precision과 Recall의 관계는 Trade-Off!**

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{All\ Detections}$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{All\ Ground\ truths}$$

		True Class	
		P	N
predicted class	P	TP	FP
	N	FN	TN

# ROC curve

## Receiver Operating Characteristic

- **Receiver Operating Characteristic(ROC) curve**

- Recall과 Fall-out의 관계를 그린 곡선임.
  - 좌상단으로 갈수록 더 좋은 커브임. (분류기의 성능이 좋다고 판단)
- **curve의 모양의 의미와 커브의 한 점의 의미는?** (뒷장에서 자세히)
  - x축: *False Positive Rate(FPR) = Fallout*
    - 실제 음성 중에서 양성으로 잘못 예측한 비율
    - $FPR = 1 - TNR = TN/(FP+TN)$
    - $TNR = specificity$  = 실제 음성 중 음성으로 정확하게 예측한 비율
  - y축: *Recall = True Positive Rate(TPR)*
    - 실제 양성 중에서 양성으로 정확하게 예측한 비율

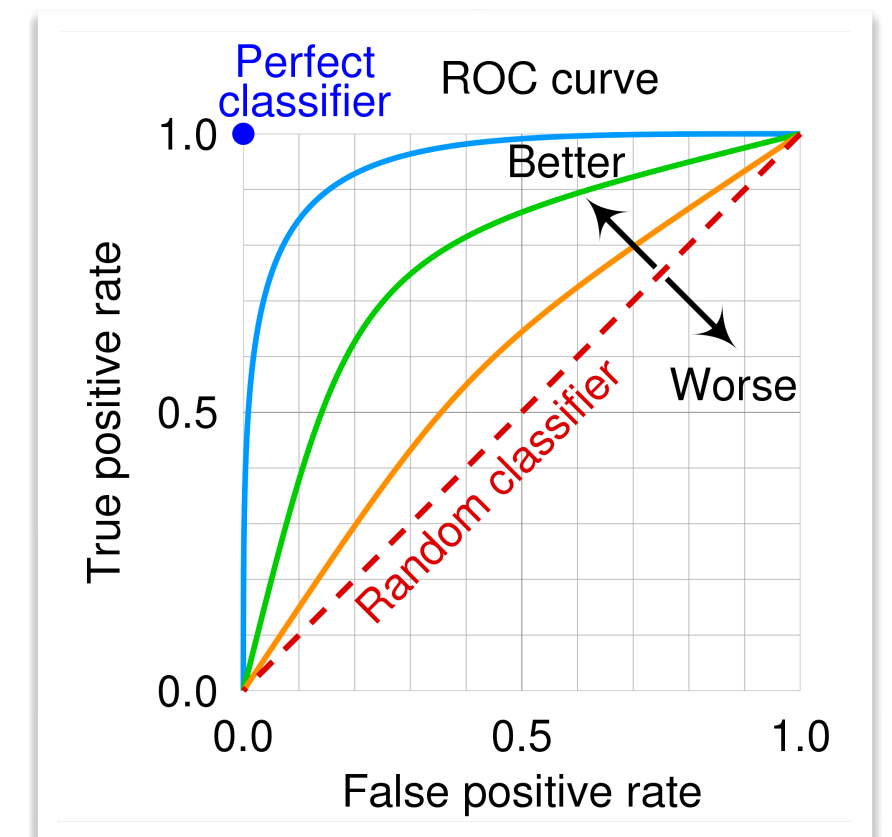
- **Area Under the Curve (AUC)**

- ROC Curve 아래의 면적을 가리킴. 최대값은 1
- 면적을 통해서도 더 좋은 커브를 찾을 수 있음.
  - AUC가 클수록 더 좋은 결과임.

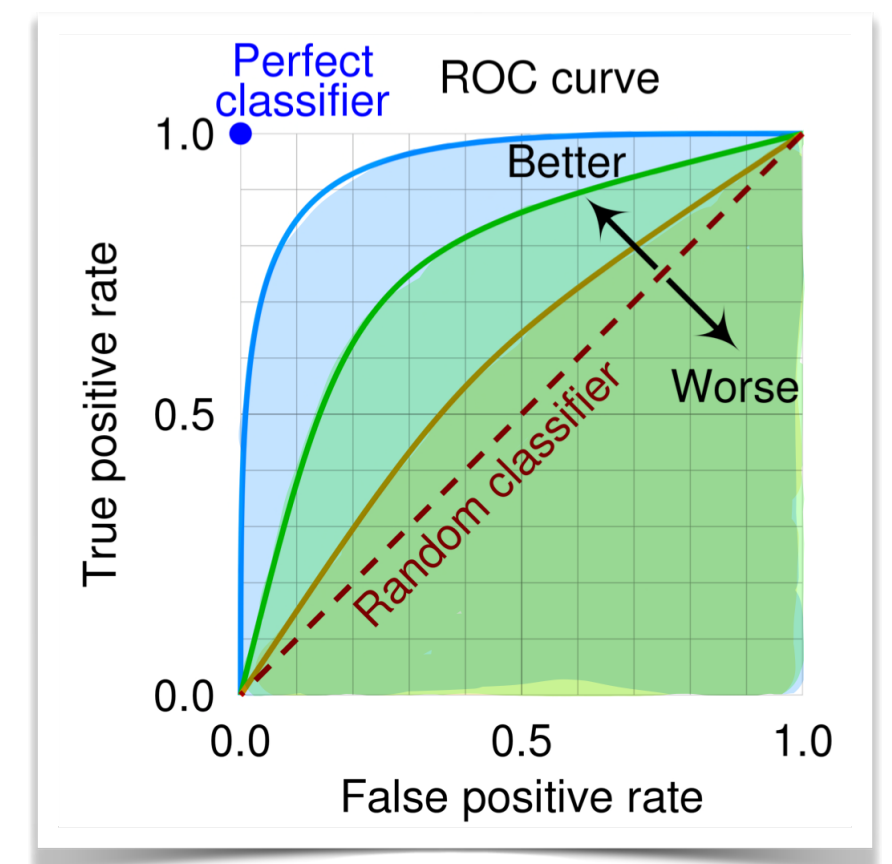
$$Fallout = \frac{FP}{FP + TN}$$

$$Recall = \frac{TP}{TP + FN}$$

ROC curve



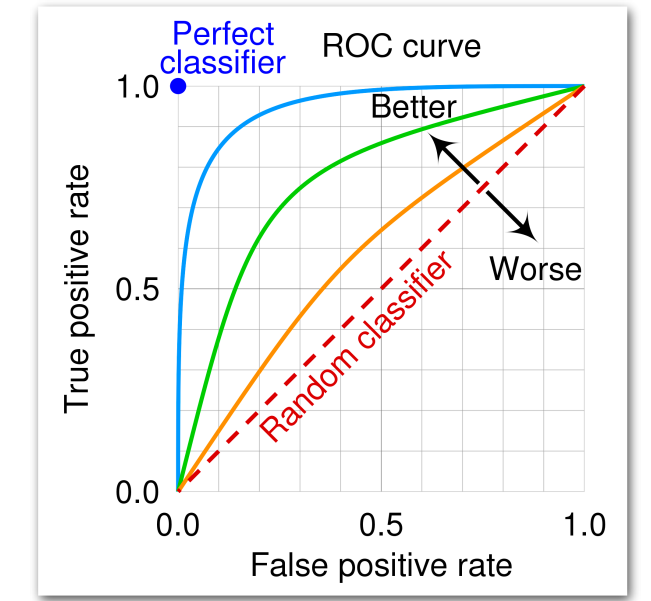
AUC



c1 > c2 > c3

# ROC curve

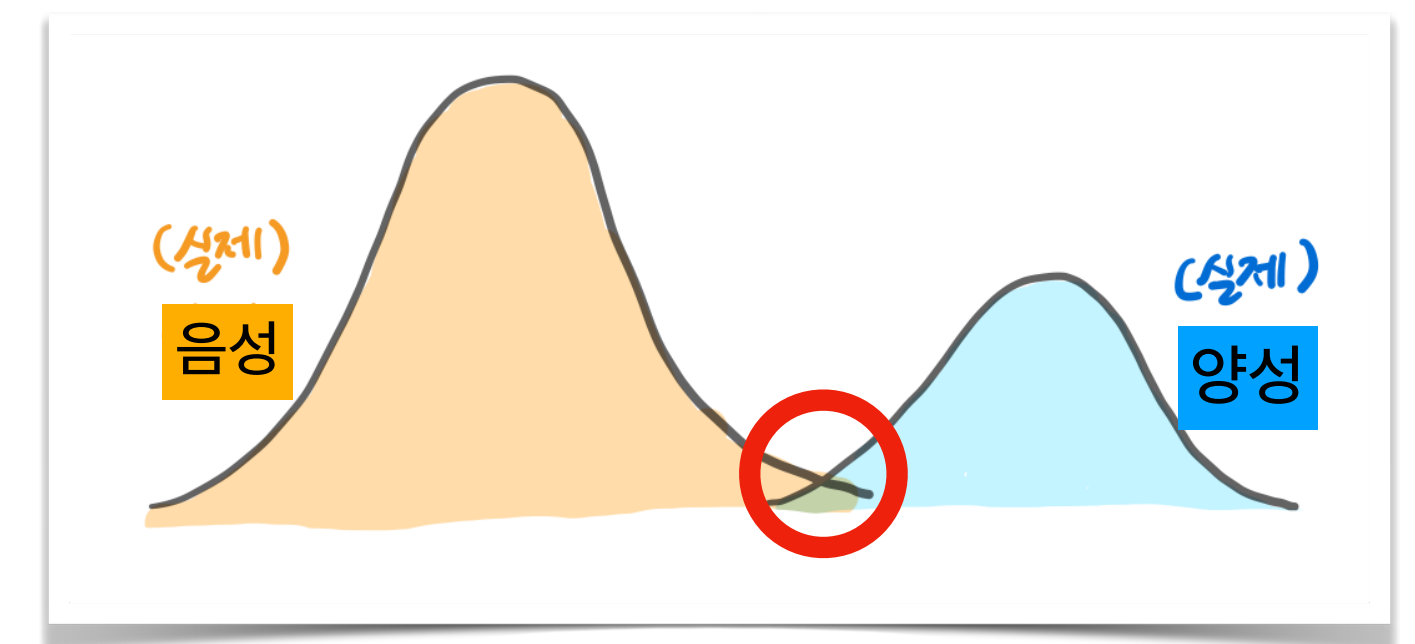
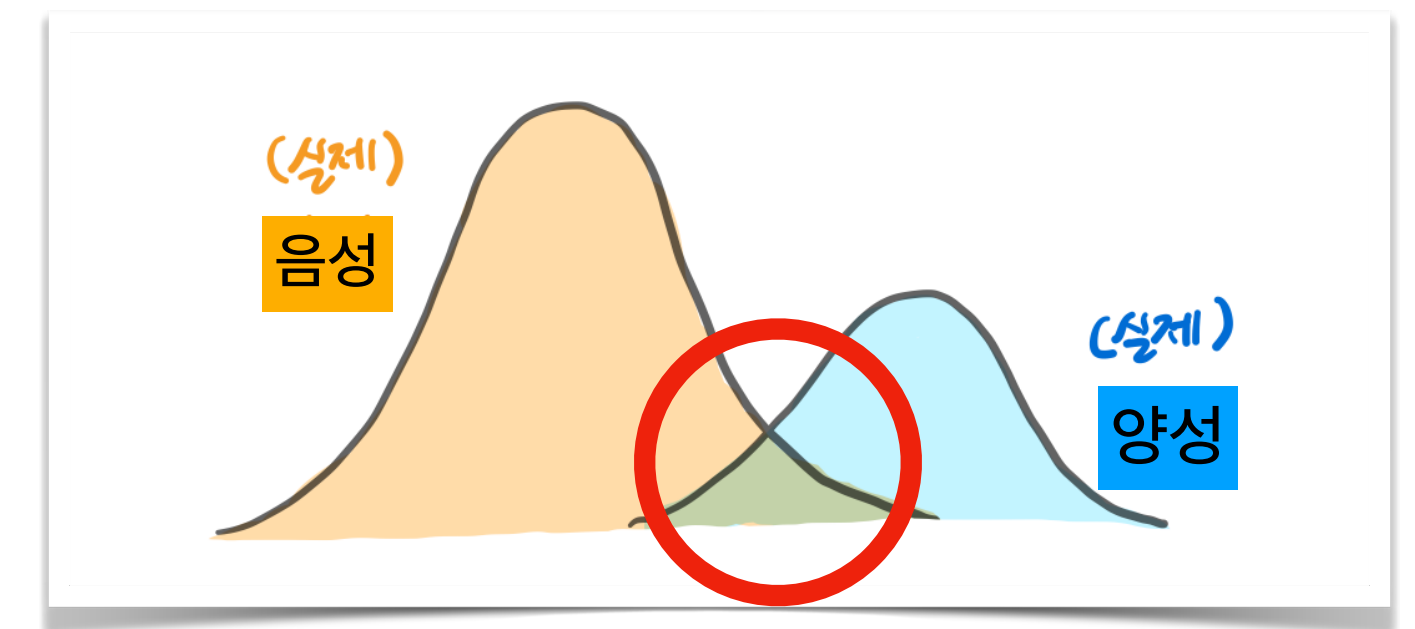
## 더 좋은 커브의 의미



### • 커브의 휨 정도의 차이

- 그래프 1과 그래프 2의 차이를 보면, 겹쳐 있는 정도가 다름.
- 겹쳐 있는 정도에 따라 커브의 휨 정도가 달라짐.
- 그림과 설명
  - 분류기가 완벽하게 분류하기 쉽지 않음. 실제 상황에서는 **불분명한 부분**이 생길 수 밖에 없음.
  - 어떤 그래프가 더 분리하기 쉬운가? = 어떤 분류기(feature)가 더 잘 판단하는가?
    - 그래프 2 > 그래프 1
    - 추론하면, 그래프 1이 c2에, 그래프 2는 c1에 가까움.
  - $TP$ (실제 양성)와  $TN$ (실제 음성)이 잘 분리되어 있을수록?
    - $AUC \approx 1$ , ROC curve는 좌상단에 가까워짐.

그래프 1(위)과 2(아래)



c1 vs c2

# ROC curve

## 커브 한 점의 의미

- **Threshold의 변화**

- 예측 기준선의 변화 (분류기 성능은 동일하지만, *sensitive*값이 달라짐)

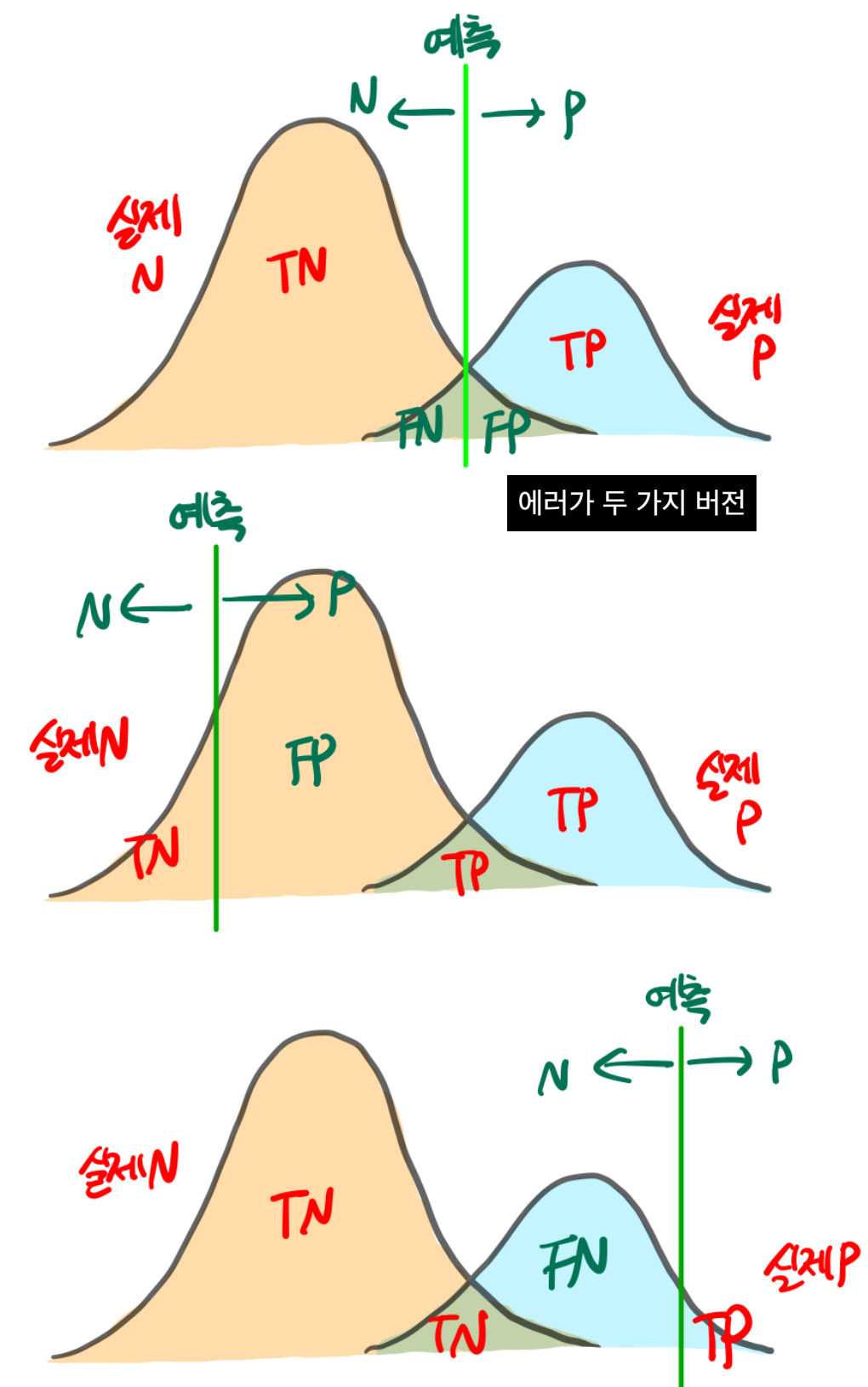
- **그림과 설명**

- *Threshold*는 초록색 기준선과 동일 (기준선 기준으로 예측 *P/N*을 나눔)
  - 그래프 1은 가장 잘 분류된 상태
  - 그래프 2는  $TPR \approx FPR$  ( $1 - TNR$ )  $\approx 1$  (보수적인 판단: 최대한 양성으로 판단)
  - 그래프 3은  $TPR \approx FPR$  ( $1 - TNR$ )  $\approx 0$  (후한 판단: 최대한 음성으로 판단)

- **의미**

- *Threshold*에 따라서 *TPR*과 *FPR*값이 동시에 바뀜.
- *Threshold*가 높아지든 낮아지든, *TPR*과 *FPR*은 비례적으로 커지거나 작아짐.
  - *TPR*과 *TNR*은 반비례
- *Threshold*의 값은 정답이 없음. 어플리케이션에 따라 개발자가 판단해야할 부분

그래프 1, 2, 3 (위-아래)



$$\frac{TP}{TP+FN} \approx 1$$
$$\frac{TN}{TN+FP} \approx 0$$
$$\Rightarrow 1 - TNR = FPR \approx 1$$

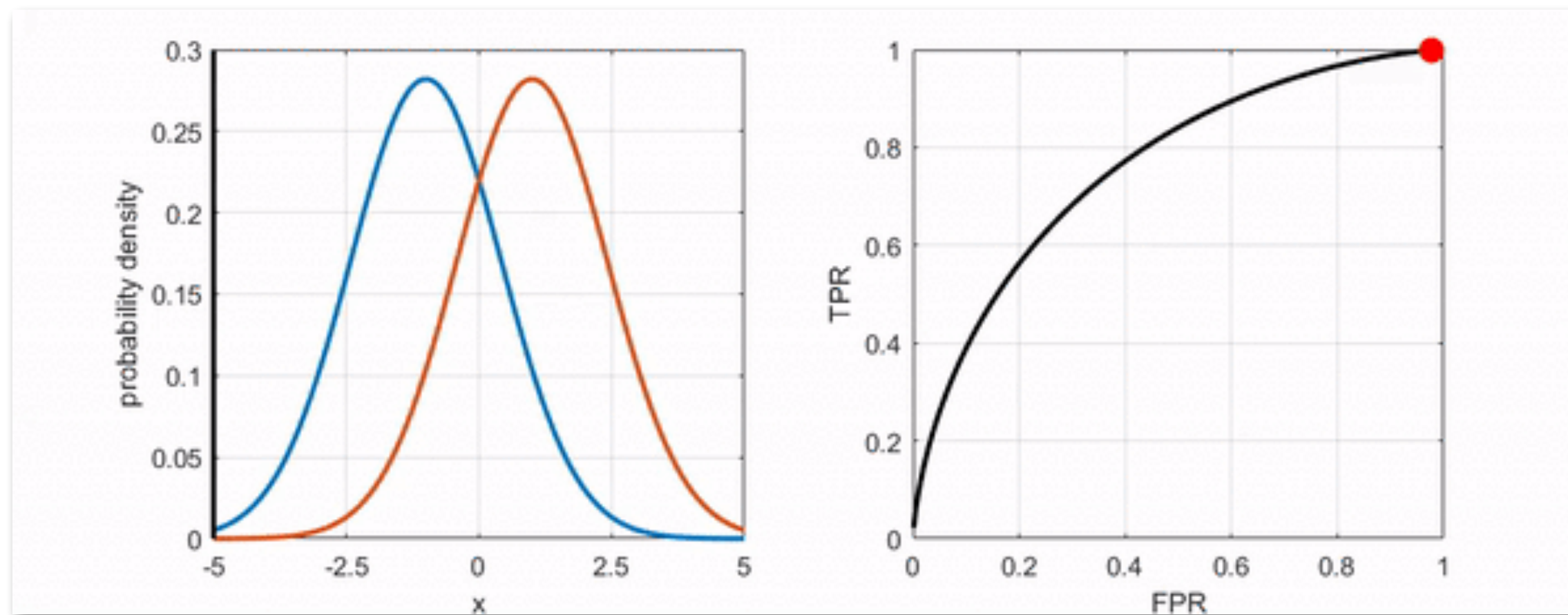
$$\frac{TP}{TP+FN} \approx 0$$
$$\frac{TN}{TN+FP} \approx 1$$
$$\Rightarrow 1 - TNR = FPR \approx 0$$



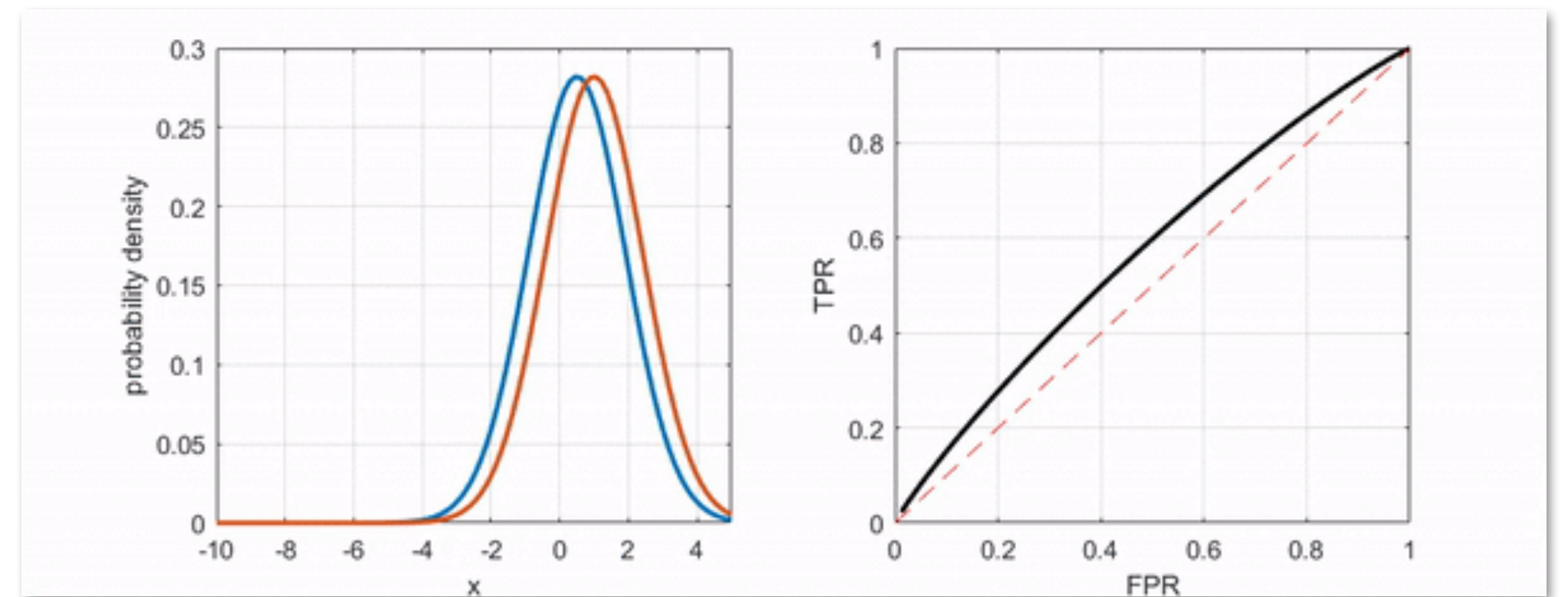
# ROC curve 정리

## 애니메이션 링크 & 보충

threshold 변화에 따른 ROC 커브 위의 점 위치 변화



분류기의 성능에 따른 ROC 커브 모양의 변화



# P-R curve

## Precision-Recall Curve

- **Precision-Recall(P-R) curve**

- *Recall*과 *Precision*의 관계를 그린 곡선임.
- **Confidence Threshold에 따른 P-R 변화**
  - Confidence(신뢰도)의 Threshold
    - Confidence: 확률이 특정 임계값 이상일 때만, 예측으로 봄. (probability >= Threshold: positive)
    - Object detection에서 확률은 IOU임.

- **그래프**

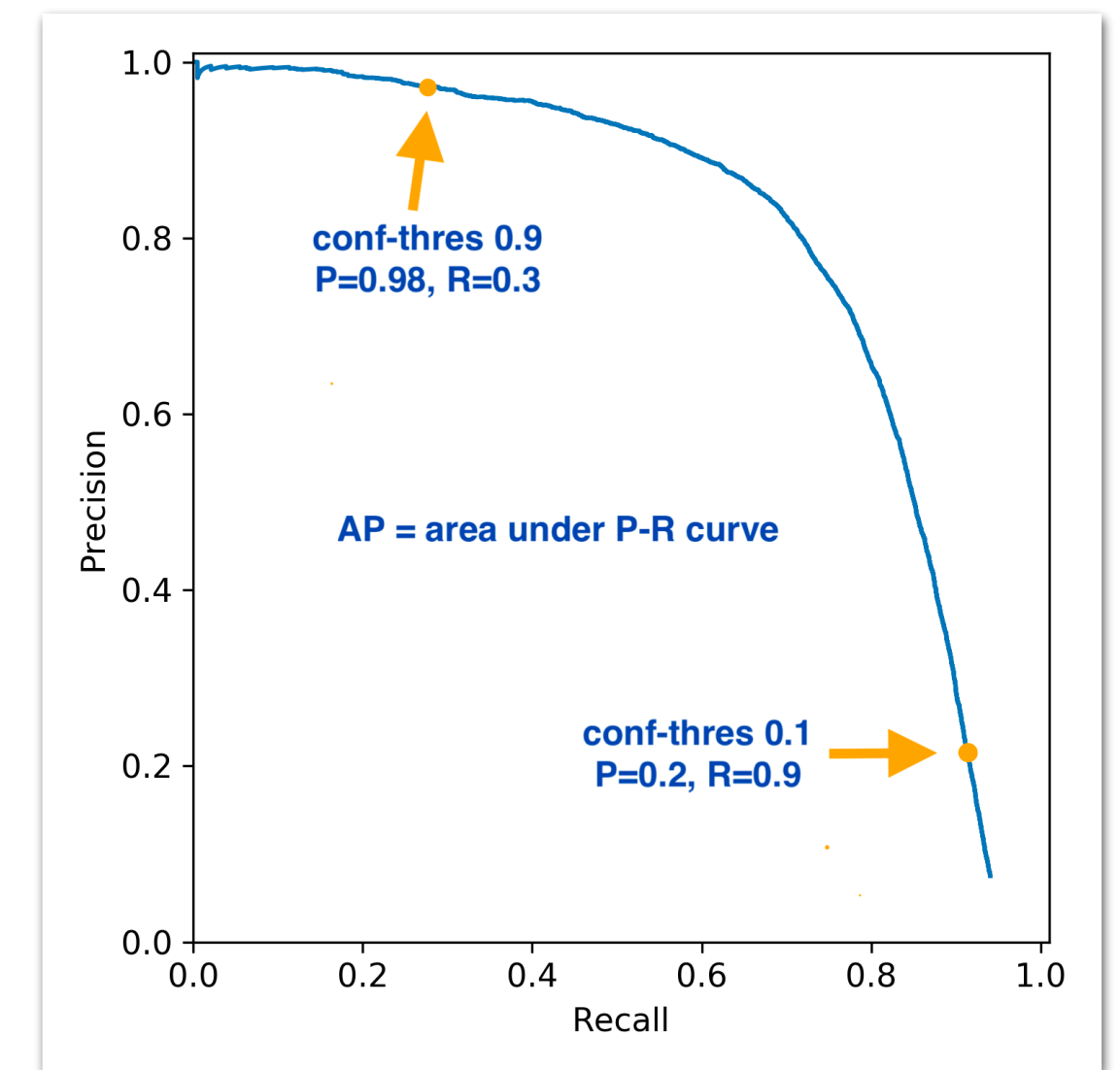
- x축: Recall, y축: Precision = *Positive Predictive Value*(PPV)
- Confidence를 조절하면서 Recall과 Precision의 변화를 그림
  - Confidence Threshold ↑ = 엄격하게 예측 = Recall 감소, Precision 증가
  - Confidence Threshold ↓ = 후하게 예측 = Recall 증가, Precision 감소

- **AP (Average Precision)**

- 커브 아래의 면적 크기 (여러개의 모델의 성능을 정량적으로 비교 가능)
- Precision, Recall 모두 높을수록 좋음 = AP가 클수록 좋음

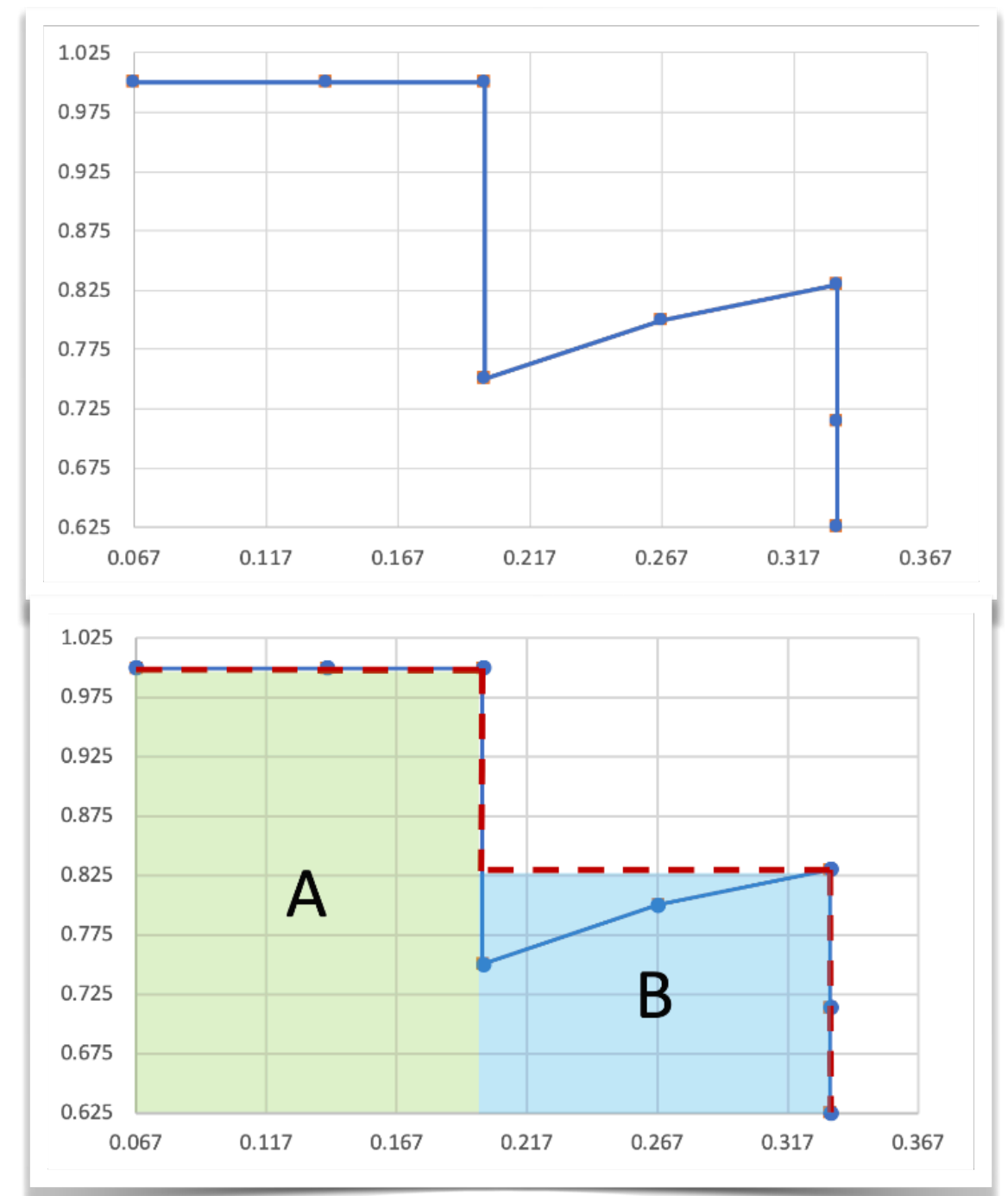
- **mAP (mean Average Precision)**

- 여러 object에 대한 AP를 평균한 값 (AP: 1개의 Object Class에 대한 수치)



$$Recall = \frac{TP}{TP + FN}$$

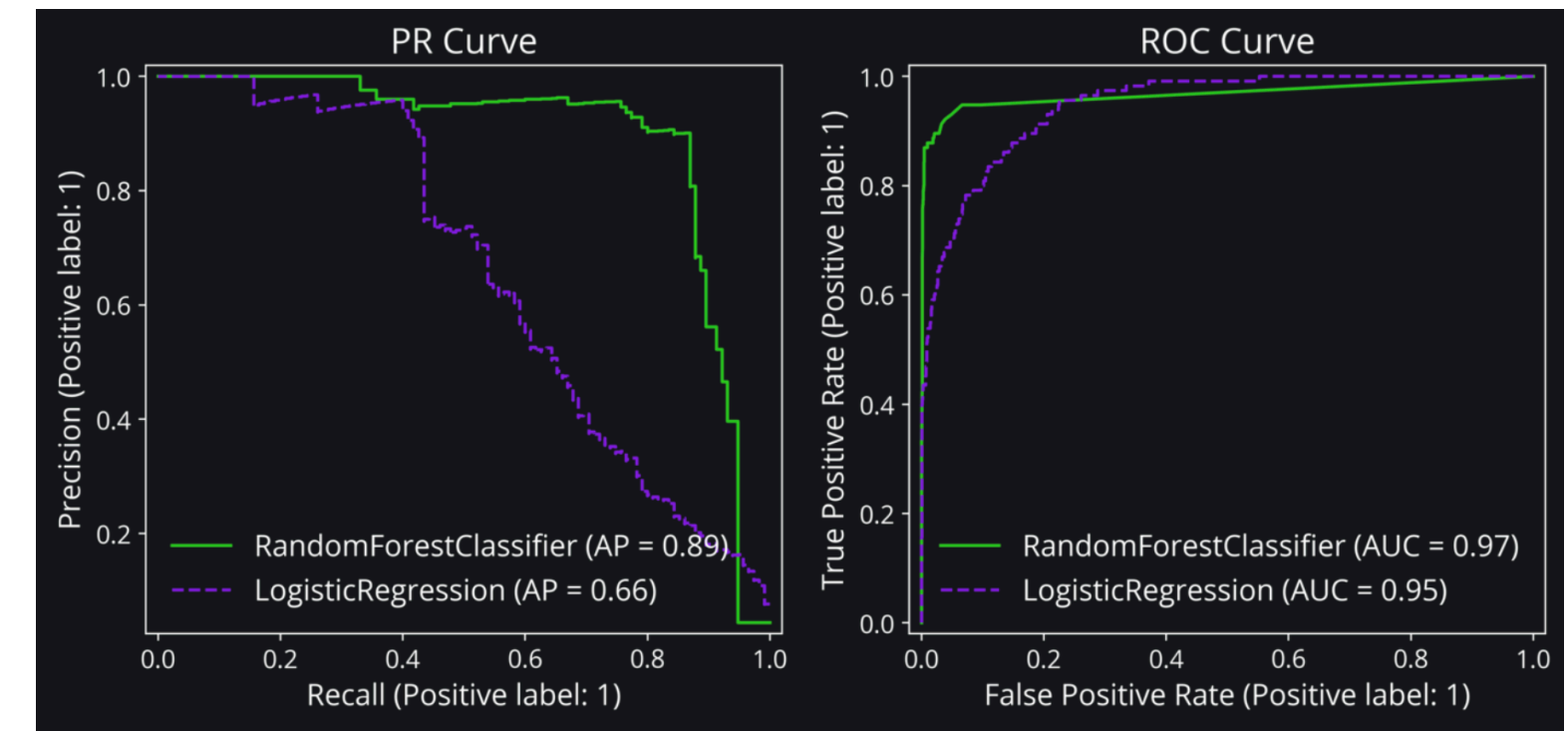
$$Precision = \frac{TP}{TP + FP}$$



$$AP = (A + B)$$

# ROC vs P-R curve

## 공통점과 차이점



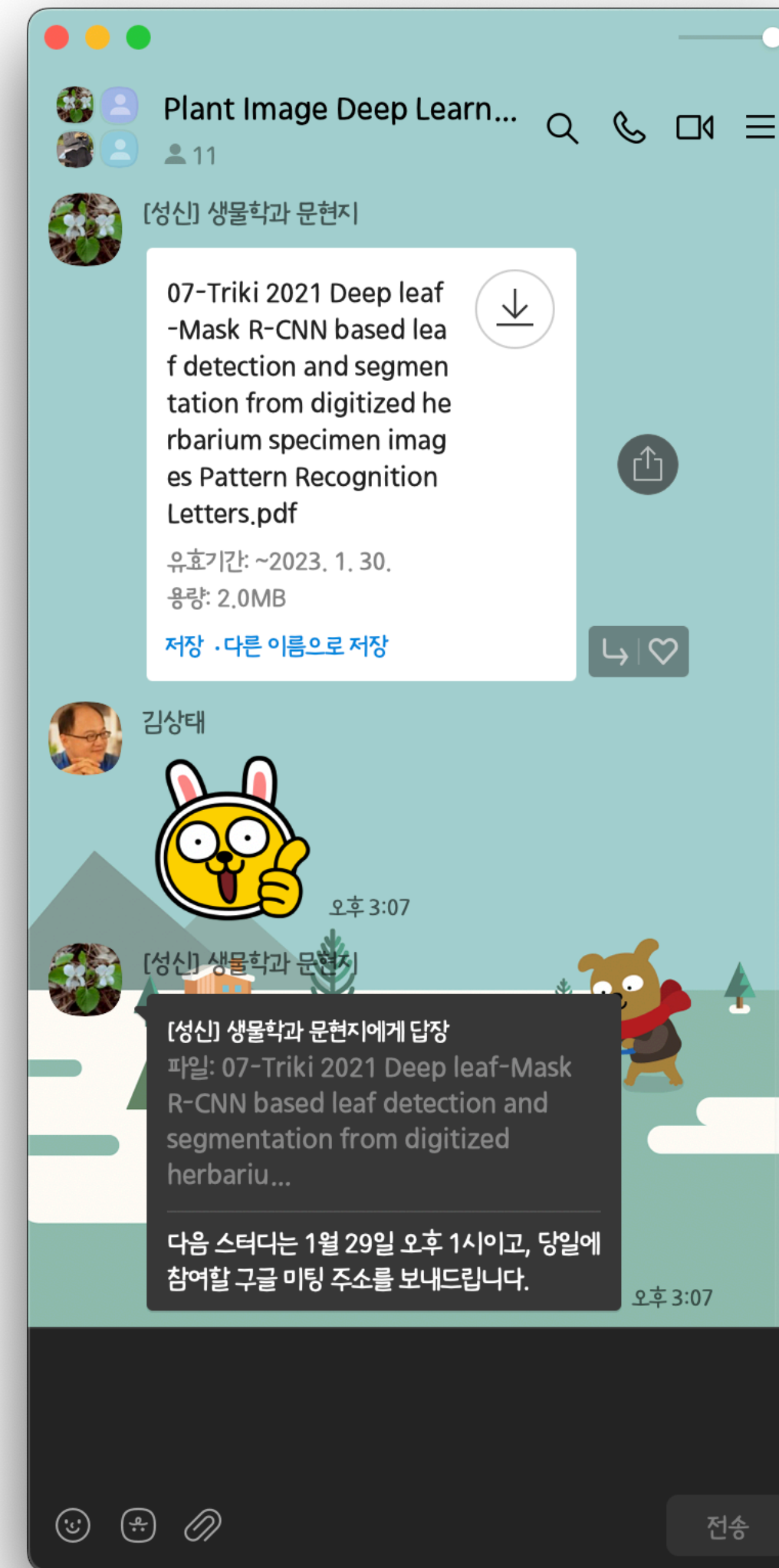
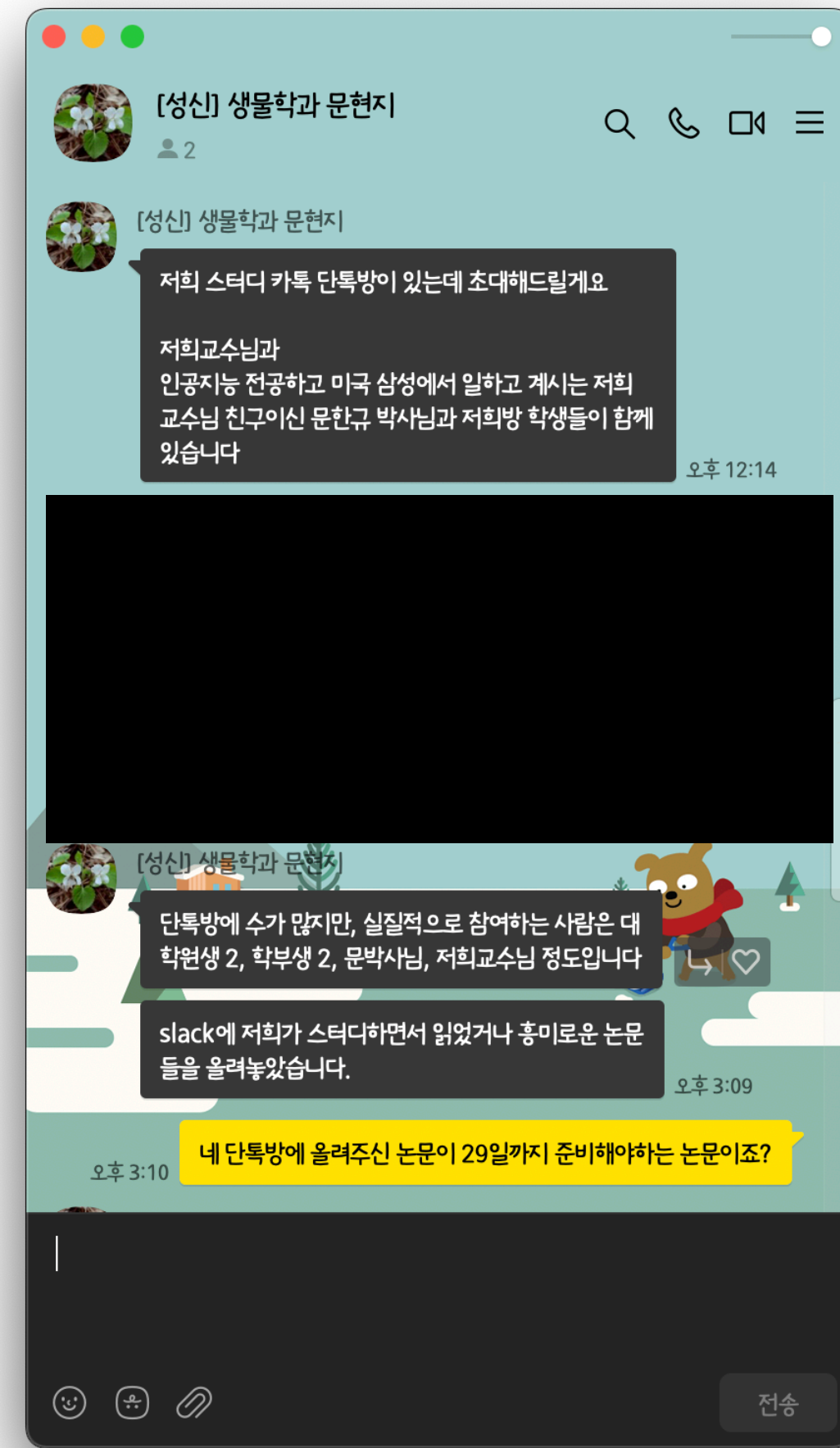
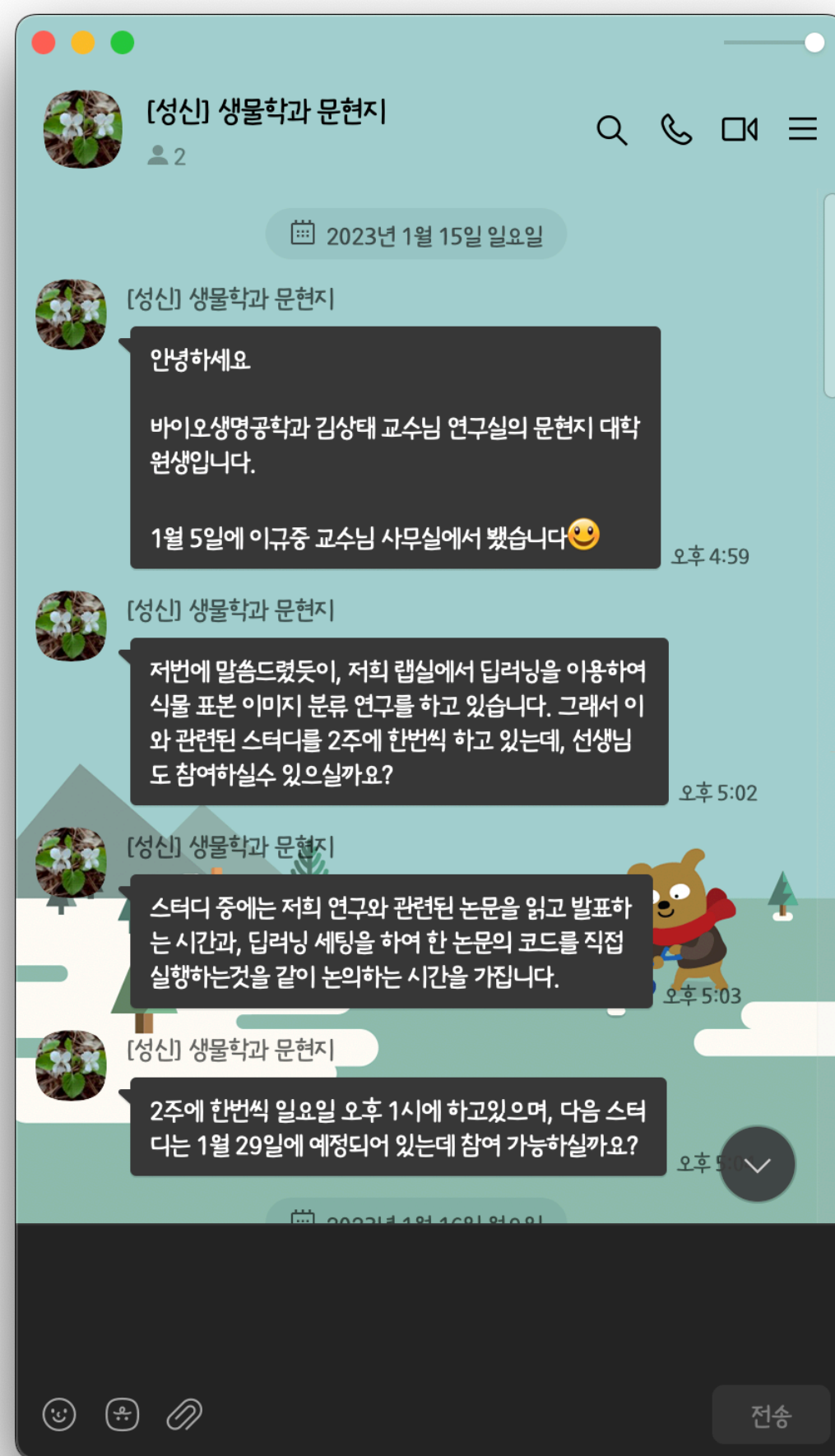
그림

- **Threshold에 따른 Precision & Recall의 변화를 나타내는 하나의 지표**
- **차이점**
  - 데이터가 불균형할 때, ROC curve보다는 P-R curve가 더 잘 반영
    - ROC curve의 TPR, FPR은 실제 양성 중 양성 예측 비율, 실제 음성 중 음성 예측 비율로 데이터의 불균형에 영향을 받지 않음.
  - 양성 클래스 탐지가 음성 클래스 탐지보다 중요도가 높을 때, ROC curve 보다는 P-R curve가 더 잘 반영
    - PR curve는 TN(실제 음성 중 예측 음성 판단)에 전혀 관여하지 않음. 음성 보다 양성을 판단하는 비중이 높음을 의미.
    - (예시) <실제로 좋아했을 영화를 추천하지 않을 수 있는 추천 시스템>보다 <이미지에 암이 있을 때 암을 감지하지 못하는 의료 영상 알고리즘>은 위음성보다 참양성이 훨씬 중요함.
- **(그림) 데이터가 불균형한 경우, PR vs ROC curve**
  - ROC curve는 일반적으로 Random Forest가 Logistic Regression보다 성능이 우수함을 보여줌. PR curve는 우리의 특정 클래스 불균형을 고려하고 두 분류기 간의 정밀도 차이가 크기 때문에 두 분류기 간의 성능 차이가 더 큼.



# 전달사항

## plant image deep learning meeting



문현지 학생 카톡)  
문현지 학생에게 연락 와서 단톡과  
Slack에 초대됐고, 스터디 참여 요청  
받았습니다.

스터디) 격주로 일요일 1시에 온라인으로 진행 (1월 29일 첫 참여)  
(저는 첫 참여라 Deep leaf: Mask R-CNN based ~ (사진 첨부)를 한번 읽고 참여를 바란다고 전달받았습니다.)