

# **평가 지표와 측정**

# 정확도

- 거짓 양성 (양성으로 판단했지만 실제 음성인 경우)과 거짓 음성 (음성으로 판단했지만 양성인 경우)의 중요도가 같지 않은 경우
  - 사례 → 암이 아닌데 암으로 판단 vs 암인데 암이 아닌 것으로 판단
- 불균형 데이터 세트
  - 90%가 양성인 데이터 세트에 대해 모두 양성으로 예측하면 정확도 90%가 도출됨
- 따라서 정확도만으로 예측 성능을 측정하는 것에는 문제가 있음

# 오차 행렬

- 이진 분류 평가 결과를 표시할 때 가장 많이 사용하는 방법

negative class	TN	FP	
positive class	FN	TP	
	predicted negative	predicted positive	

# 오차 행렬

- 비슷한 정확도를 보이는 모델을 오차 행렬로 표현

모두 음성으로 판단한 모델:

$\begin{bmatrix} 403 & 0 \\ 47 & 0 \end{bmatrix}$  정확도 : 90%

무작위 더미 모델:

$\begin{bmatrix} 357 & 46 \\ 43 & 4 \end{bmatrix}$  정확도 : 82%

결정 트리:

$\begin{bmatrix} 390 & 13 \\ 24 & 23 \end{bmatrix}$  정확도 : 92%

로지스틱 회귀

$\begin{bmatrix} 401 & 2 \\ 8 & 39 \end{bmatrix}$  정확도 : 98%

# 정확도, 정밀도, 재현율, F-Score

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{정밀도} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{재현율} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F = 2 \cdot \frac{\text{정밀도} \cdot \text{재현율}}{\text{정밀도} + \text{재현율}}$$

- 정밀도 또는 재현율 하나만으로는 전체 그림을 볼 수 없기 때문에 이 둘의 조화 평균인 F-Score 사용

# 오차 행렬

- 비슷한 정확도를 보이는 모델을 오차 행렬로 표현

모두 음성으로 판단한 모델:

```
[[403  0]  
 [ 47  0]]
```

정확도 : 90%

F-Score : 0.00

무작위 더미 모델:

```
[[357 46]  
 [ 43  4]]
```

정확도 : 82%

F-Score : 0.08

결정 트리:

```
[[390 13]  
 [ 24 23]]
```

정확도 : 92%

F-Score : 0.55

로지스틱 회귀

```
[[401  2]  
 [  8 39]]
```

정확도 : 98%

F-Score : 0.89

- classification\_report 함수를 사용해서 요약된 정보 표시

# 불확실성 고려

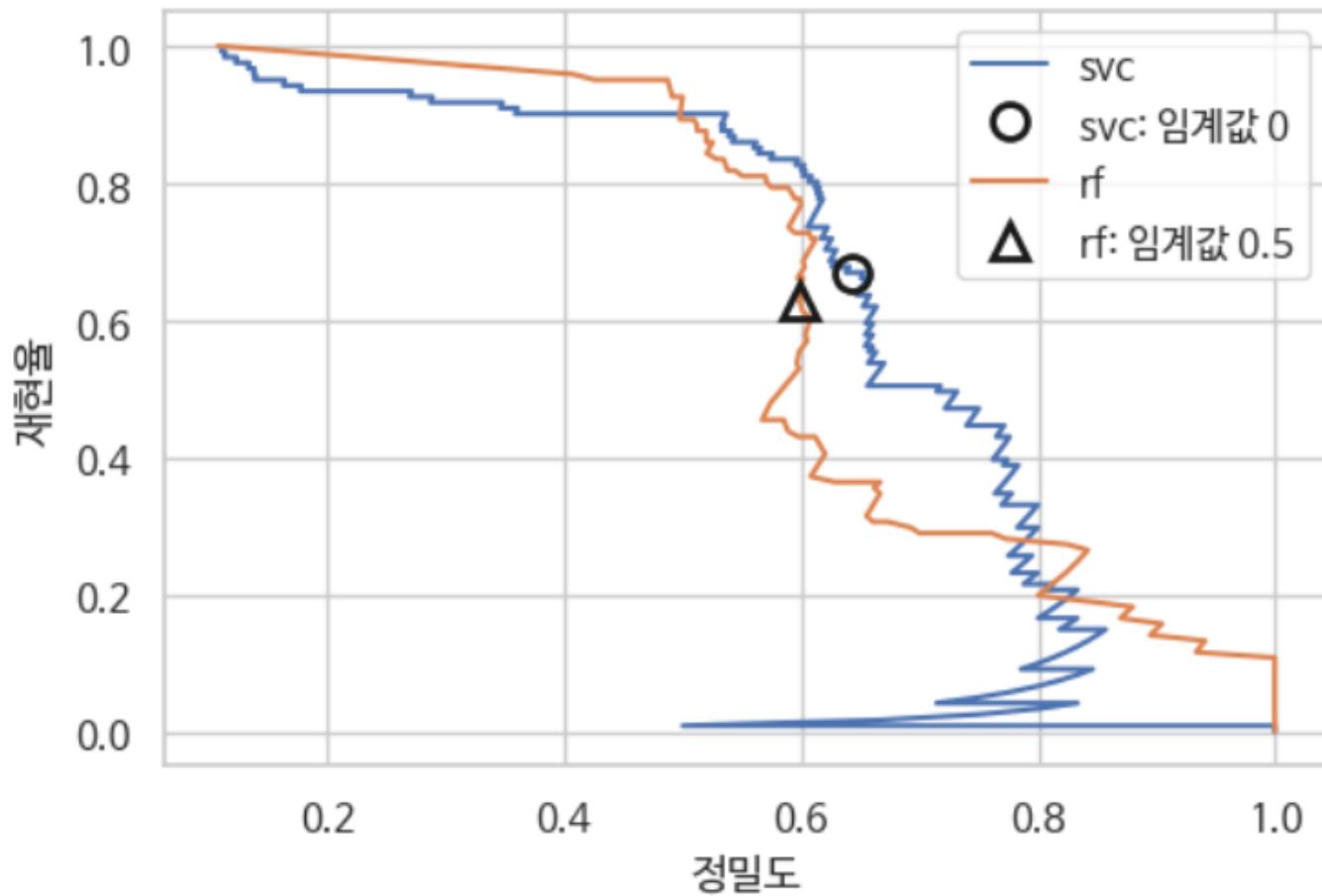
- 대부분의 분류기는 예측의 확신을 가늠하기 위해 `decision_function` 또는 `predict_proba` 함수 제공
  - `decision_function`의 임계값은 0
  - `predict_proba`의 임계값은 0.5
- 임계점을 조정해서 재현율에 더 가중치를 두거나 혹은 정밀도에 더 가중치를 두도록 구현할 수 있음

# 정밀도-재현율 곡선

- 모델의 분류 작업을 결정하는 임계값을 변경하는 것은 해당 모델의 정밀도와 재현율의 상충 관계를 조정하는 것
- 적절한 임계값 설정을 위해 정밀도와 재현율을 종합적으로 살펴보아야 하며 이를 위해 정밀도-재현율 곡선 사용
- scikit-learn의 `precision_recall_curve` 함수 사용해서 정밀도-재현율 곡선 표시
- 특정 임계값이나 운영 포인트에 국한하지 않고 전체 곡선에 담긴 정보를 요약하기 위해 정밀도-재현율 곡선의 아래 부분 면적을 계산
  - 평균 정밀도
  - `average_precision_score` 함수 사용

# 정밀도 - 재현율 곡선

- 정밀도 - 재현율 곡선 사례

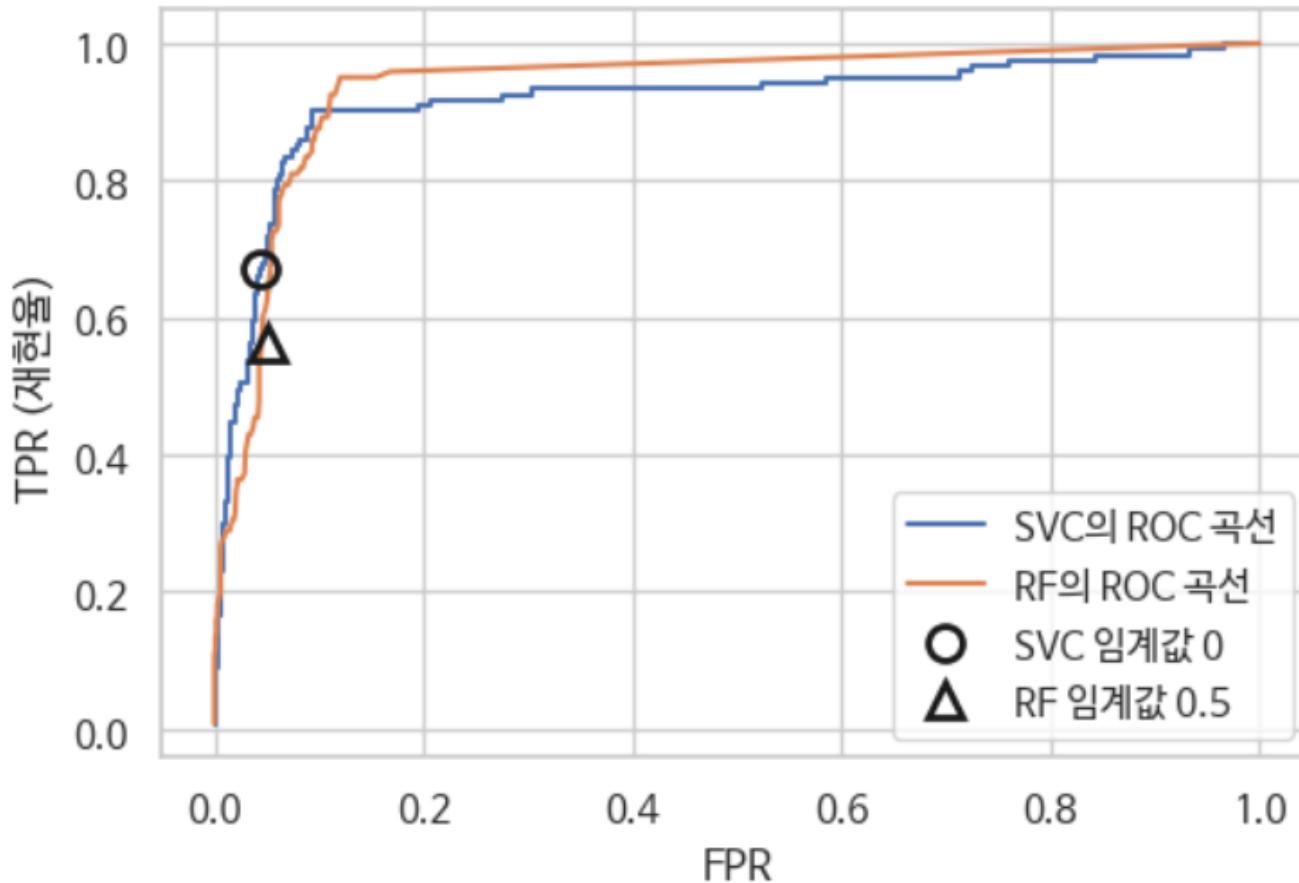


# ROC 곡선

- 여러 임계값에서 분류기의 특성을 분석하는데 널리 사용되는 도구
- 정밀도 - 재현율 대신 진짜 양성 비율에 대한 거짓 양성 비율을 표현
  - 진짜 양성비율은 재현율 (TPR)
  - 거짓 양성 비율은 전체 음성 샘플 중 양성으로 잘못 분류한 비율 (FPR)
- `roc_curve` 함수를 사용해서 표시
  - 좌측 상단에 가까울수록 이상적

# ROC 곡선

## ▪ ROC 곡선 사례



# AUC

- ROC 곡선 아래쪽의 면적 값으로 ROC 곡선을 요약할 때 사용
- `roc_auc_score` 함수로 구현