

- 머신러닝에서 어떠한 문제를 해결하는 시스템을 모델이라고 부름
- 문제의 유형에 따라 성능을 측정할 수 있는 지표가 다름

Regression

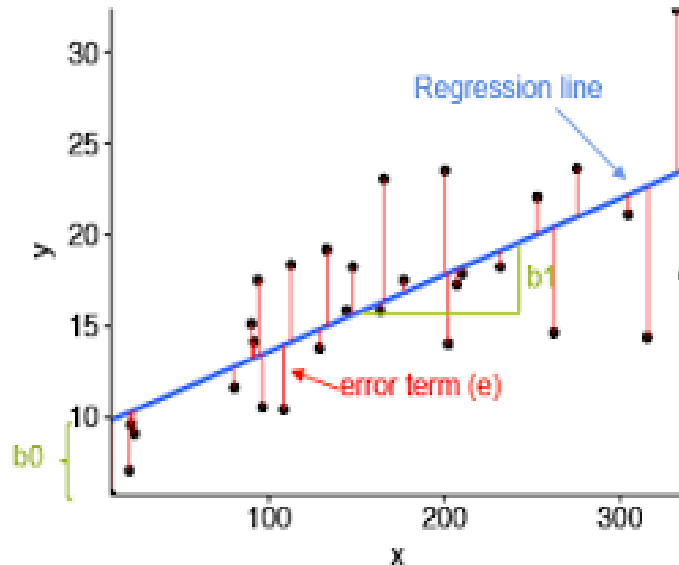
- MSPE
- R Square
- Adjusted R Square

Classification

- Precision-Recall
- ROC-AUC
- Accuracy
- Log-Loss

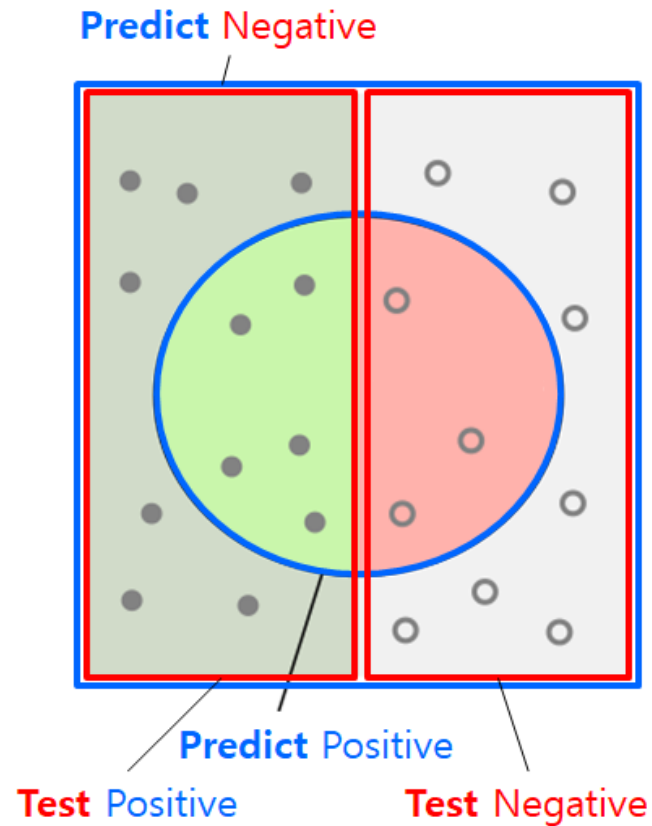
- Regression(회귀)

- MSPE: 정답과 예측 값 차의 제곱 가중합
- R Square : MSE / 데이터 셋의 분산
- Adjusted R square : R Square 값에 input feature의 개수 패널티 가중치를 준 값



$$MSPE = \frac{100\%}{N} \sum_{i=1}^N \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2$$

- Classification(분류)



- **Test** 데이터 셋

- Positive / Negative

- **Predict** 결과 셋

- Positive / Negative

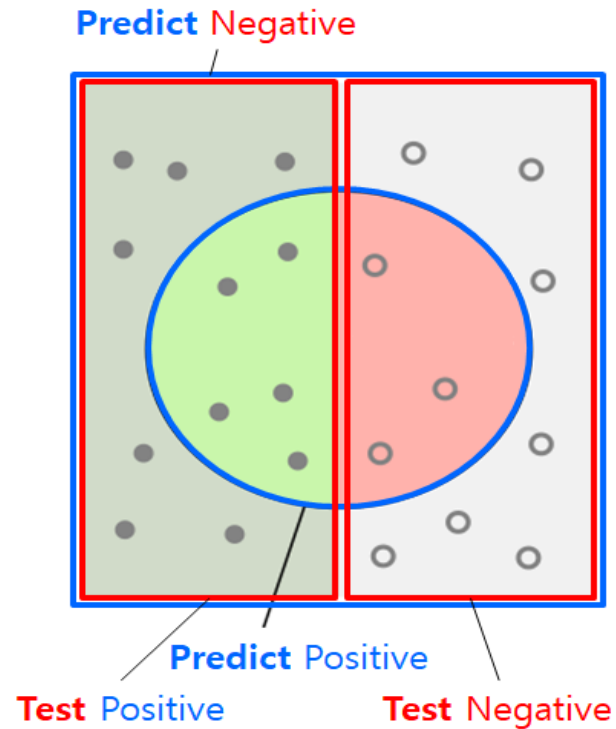
- **True:**

- Predict / **Test** 일치

- **False:**

- Predict / **Test** 불일치

- Classification(분류)
 - Confusion Matrix



		실제 정답	
		Positive	Negative
실험 결과	Positive	True Positive	False Positive (Type II error)
	Negative	False Negative (Type I error)	True Negative

- Classification(분류)

- Accuracy : 정합도, 모델이 정답을 정답으로 오답을 오답으로 선택하는 가를 보여주는 지표

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

		실제 정답	
		Positive	Negative
실험 결과	Positive	True Positive	False Positive (Type II error)
	Negative	False Negative (Type I error)	True Negative

- Classification(분류)

- 1000명 중 10명(1%) 존재하는 암환자를 진단하는 모델 구축

	실제	암진단모델
정상	990	999
암	10	1

- 9명은 자신이 암에 걸린 지 모르고 치료시기를 놓침 => **사망**
 - 정확도는 **99.1%**(991/1000) 이지만 좋은 '암진단모델' 이라고 말할 수 있을까?

- Classification(분류)

- Precision : 정밀도, 모델에서 정답이 아닌 것을 얼마나 잘 걸러 내는가를 보여주는 지표

$$\text{Precision} = \frac{tp}{tp + fp}$$

		실제 정답	
		Positive	Negative
실험 결과	Positive	True Positive	False Positive (Type II error)
	Negative	False Negative (Type I error)	True Negative

- Classification(분류)

- Recall : 재현율, 모델에서 실제 정답을 얼마나 많이 선택하는 가를 보여주는 지표

$$\text{Recall} = \frac{tp}{tp + fn}$$

		실제 정답	
		Positive	Negative
실험 결과	Positive	True Positive	False Positive (Type II error)
	Negative	False Negative (Type I error)	True Negative

- Classification(분류)

- 1000명 중 10명(1%) 존재하는 암환자를 진단하는 모델 구축

진단결과		실제	
		암	정상
모델	암	1	0
	정상	9	990

Accuracy	$(990 + 1) / (990 + 9 + 1 + 0)$	99.1%
Precision	$(1) / (1 + 0)$	100%
Recall	$(1) / (9 + 1)$	10%

- **Recall이 중요한 경우**

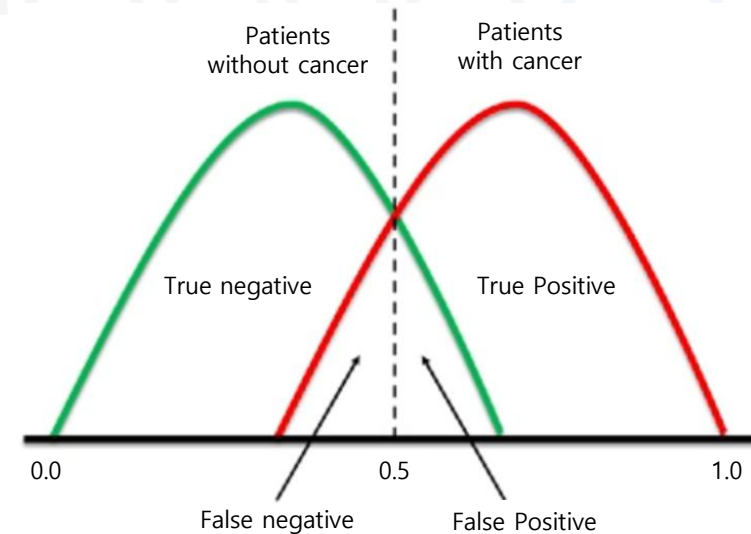
- Test 데이터가 Positive 인 데이터를 Negative로 잘못 하게 되면 업무상 큰 영향이 발생하는 경우
- Classification의 경우 대부분이 검출하고자 하는 경우가 Minor
- 대부분의 케이스에서 Precision 보다 중요한 지표
- ex) 보험사기 검출, 금융사기 검출, 어뷰징 검출, 질병 검출, Etc.

- Precision이 중요한 경우

- Test 데이터가 Negative인 데이터를 Positive로 잘못 하게 되면 업무상 큰 영향이 발생하는 경우
- ex) 스팸 메일 분류, 추천시스템, Etc.

- Classification(분류)

- 업무 특성상 Precision / Recall 중 하나의 값이 중요할 경우 Threshold (default: 0.5)를 조정
- Threshold 값 조정 시 Precision / Recall 값은 Trade-off 관계로 변화됨
- Threshold 는 업무 환경에 맞도록 모델을 튜닝하는 선에서 조정
- 과하게 조절 시 성능 지표가 무의미해짐



- Classification(분류)

- 각 각의 성능 지표가 높은 모델을 2 stage 시스템을 구축 하기도 함

