

모델평가와 성능개선

5기 이세린

CONTENTS

1. 모델평가

- 회귀모델의 평가지표
- 분류모델의 평가지표

2. 성능개선

- Cross-validation
- Grid search

10/12/2019

1. 모델 평가

3

모델평가의 목적

1. 모델의 일반화 정확도 추정

↳ 미래의 처음 본 데이터에 대한 예측성능

2. 학습 알고리즘을 튜닝하고 주어진 가설 공간 안에서 가장 성능 좋은 모델을 골라 예측 성능 향상

3. 최대 성능을 내는 알고리즘을 선택하기 위해 여러 알고리즘들 비교

모델평가지표 - Regression

- MSE

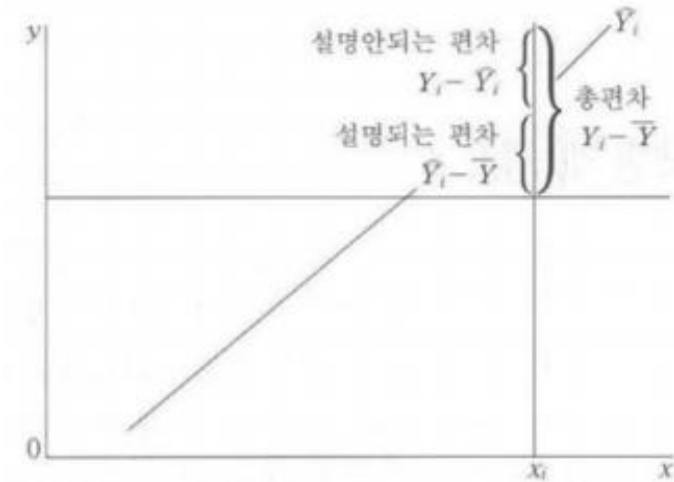
$$MSE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}$$

- RSE

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

- R²

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$



$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})}{\sum(y_i - \bar{y})^2} = \frac{\text{회귀선에 의해 설명되는 변동}}{\text{전체 변동}}$$

$$\text{adjusted } R^2 = 1 - \frac{n-1}{(n-p-1)(1-R^2)}$$

모델평가지표 - Classification

오차행렬

[Confusion Matrix]

		Predicted Data		Total
		Predicted Condition POSITIVE	Predicted Condition NEGATIVE	
Actual Data	Condition TRUE	TP True Positive	FN False Negative	P
	Condition FALSE	FP False Positive	TN True Negative	N
Total		P^+	N^+	P+N

True Positive (TP): 양성을 양성으로 제대로 검출

True Negative (TN): 음성을 음성으로 제대로 검출

False Negative (FN): 양성을 음성으로 잘못 검출

False Positive (FP): 음성을 양성으로 잘못 검출

모델평가지표 - Classification

1. 정확도(Accuracy)

- 실제값과 예측값이 일치하는 비율

$$\text{accuracy} = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + FN + FP + TN}$$

		Predicted Data		Total
		Predicted Condition POSITIVE	Predicted Condition NEGATIVE	
Actual Data	Condition TRUE	TP True Positive	FN False Negative	P
	Condition FALSE	FP False Positive	TN True Negative	N
Total		P`	N`	P+N

- 직관적으로 모델 예측 성능을 나타내는 평가지표
- But, 비대칭한 데이터 세트에서 분류의 정확도가 매우 높게 나는 수치적인 판단 오류를 일으키게 됨
- `accuracy_score()`

모델평가지표 - Classification

2. 정밀도(Precision)

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

- 예측값이 Positive인 대상 중 실제값도 Positive인 데이터의 비율
- Ex) 스팸메일 분류
- `precision_score()`

3. 재현율(Recall) = 민감도(Sensitivity)

$$\text{sensitivity} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

- 실제값이 Positive인 대상 중 예측값도 Positive인 데이터의 비율
- Ex) 암 판단 모델, 금융사기
- `recall_score()`

모델평가지표 - Classification

4. F1 score

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$F_{\beta} = \frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$$

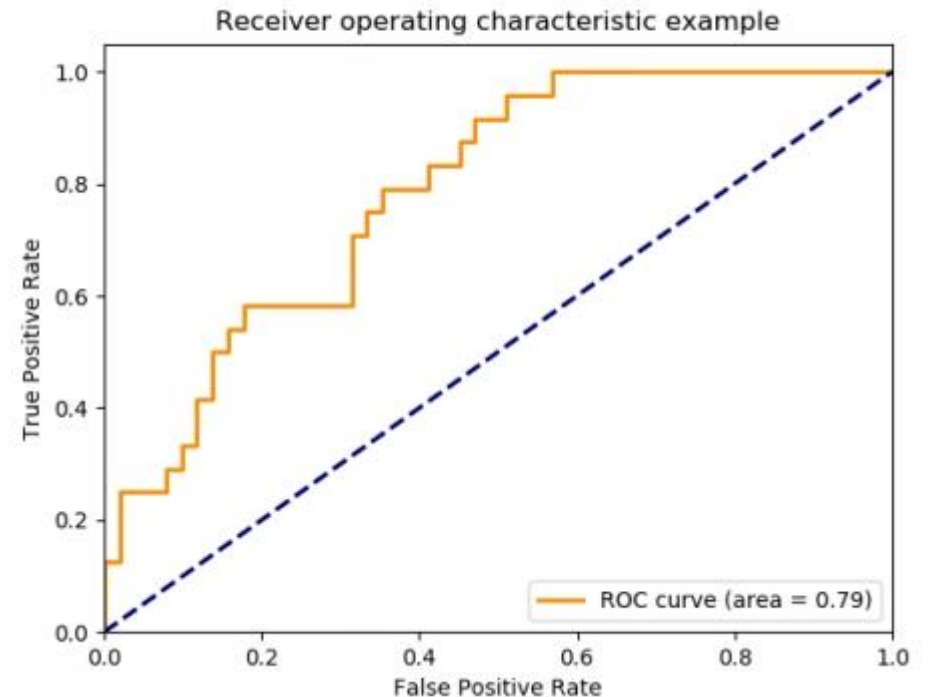
- 정밀도와 재현율의 조화평균
- 정밀도와 재현율이 어느 한쪽으로 치우치지 않는 수치를 나타낼 때 상대적으로 높은 값을 가짐(trade-off)
- F_{β} score: 정밀도의 β 배 만큼 재현율에 가중치
- `f1_score()`

ROC곡선 (Receiver Operating Characteristics)

- Specificity 를 x 축 , Sensitivity 를 y 축으로 나타낸 그래프

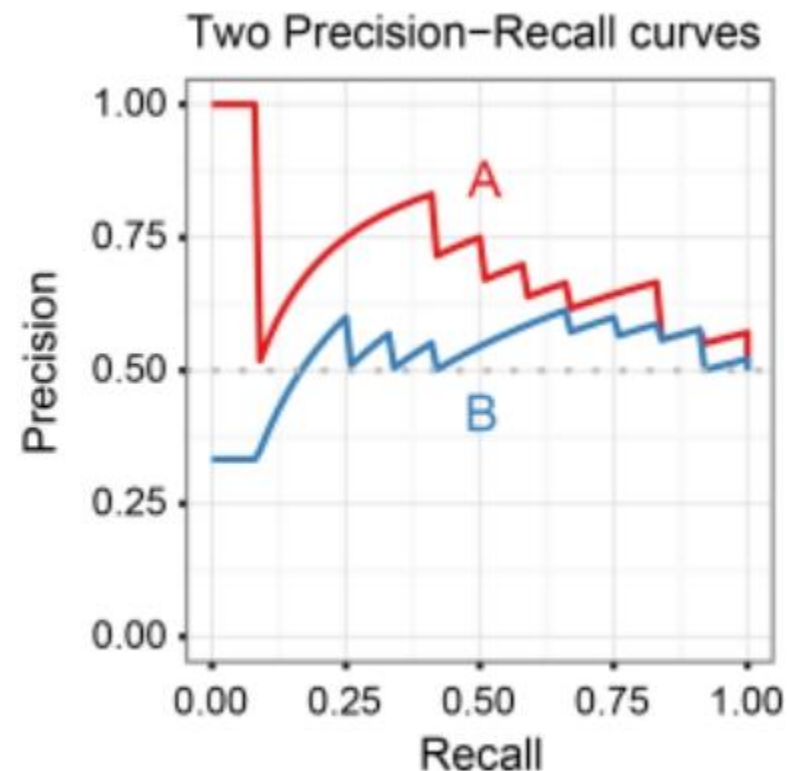
$$\text{specificity} = \frac{TN}{N} = \frac{TN}{FP + TN}$$

- 그래프가 왼쪽 위 코너로 붙을수록 좋은 모델
- $y=x$ 보다 위에 있어야 확률이 0.5 를 넘겨서 쓸모가 있는 모델
- AUC(Area Under Curve) 값으로 평가
 - ROC 그래프의 면적
 - 1 에 가까울 수록 좋음



PR(Precision-Recall) 곡선

- Recall을 x축, Precision을 y축으로 나타낸 그래프
- 데이터 라벨의 분포가 불균등할 때 사용
ex) 정상거래 98%, 비정상거래 2%
- 그래프가 오른쪽 위 코너로 붙을수록 좋은 모델
- Base line = $P/(P+N)$ 보다 위에 있어야 쓸모 있는 모델
(N: 전체데이터)
- AUC값으로 모델의 정확도 평가



실습 & 영상 😊

0.0 다중 분류 모델 성능 측정

<https://www.youtube.com/watch?v=8DbC39cvvis&t=57s>

0.1 ROC 곡선의 이해

<https://www.youtube.com/watch?v=nMAtFhamoRY&t=61s>

1. Confusion matrix 실습
2. ROC, PR 곡선 실습
3. Cross Validation 실습
4. Grid Search 실습

실습 & 영상 😊

0.0 다중 분류 모델 성능 측정

0.1 ROC 곡선의 이해

1. Confusion matrix 실습

2. ROC, PR 곡선 실습

3. Cross Validation 실습

4. Grid Search 실습

10/12/2019

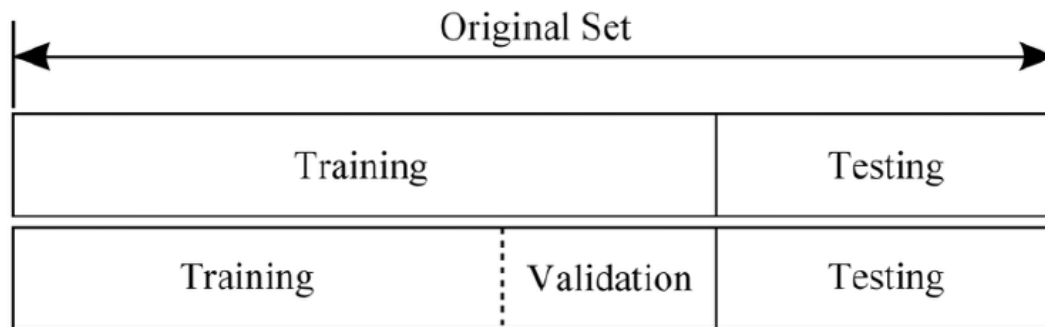
2. 성능개선

14

교차검증(Cross-Validation)

- 통상적 정확도 검정 방식

- i) Train set와 Test set을 나누어서 (일반적으로 7:3)
- ii) Train set에 기반해 모델을 만들고
- iii) Test set에 대한 예측력 및 정확도를 도출

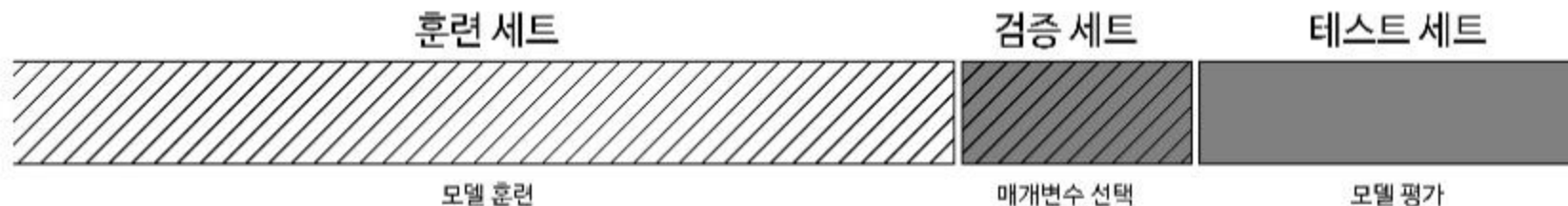


[문제점] 특정 Test set에만 최적의 성능을 발휘하는 편향된 모델 유도

Validation Set

학습용 데이터를 학습용(test)과 검증용(validation)으로 분리해
학습용 데이터로 학습한 후 검증용 데이터로 검증

& 최종 평가를 위해서는 독립된 데이터 세트가 필요 (일반화)

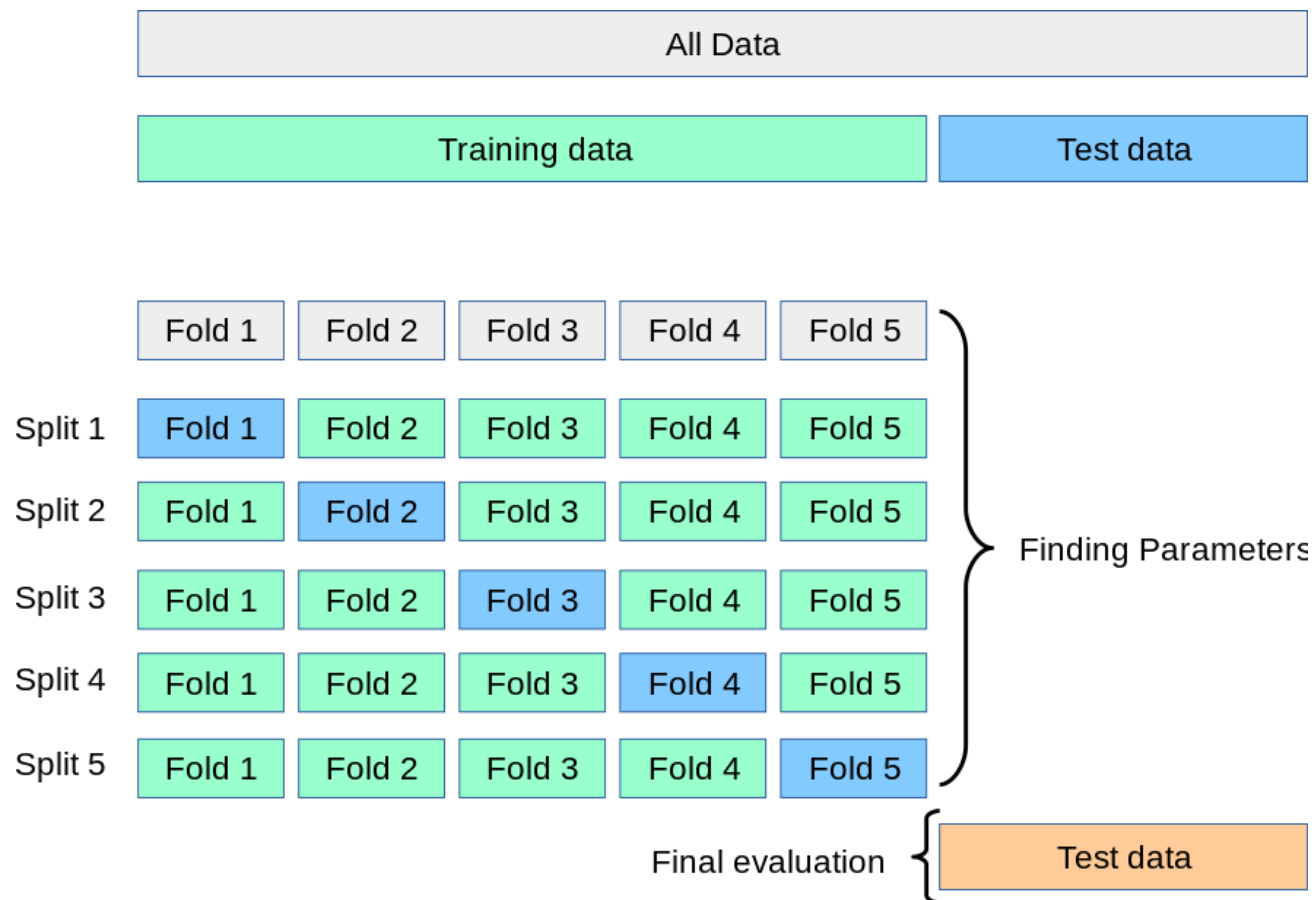


교차검증(Cross-Validation)

- 데이터를 어떻게 분리하느냐에 따라 검증 성능이 달라질 수 있음
→여러가지 방식으로 데이터를 분리하여 검증을 실시하는 방식
- 평균 성능(mean performance)& 성능분산(performance variance)
- 데이터의 수가 적은 경우, 검증 성능의 신뢰도가 떨어지는 문제 해결

K-fold Cross Validation

1. 데이터를 K개 부분집합(fold)로 분리
2. 첫번째 fold를 test set으로 두고, 나머지 K-1개 폴드를 training set으로 이용하여 K개 모델을 추정
3. 테스트 폴드를 바꾸어 가며 모든 폴드가 사용될 때까지 반복

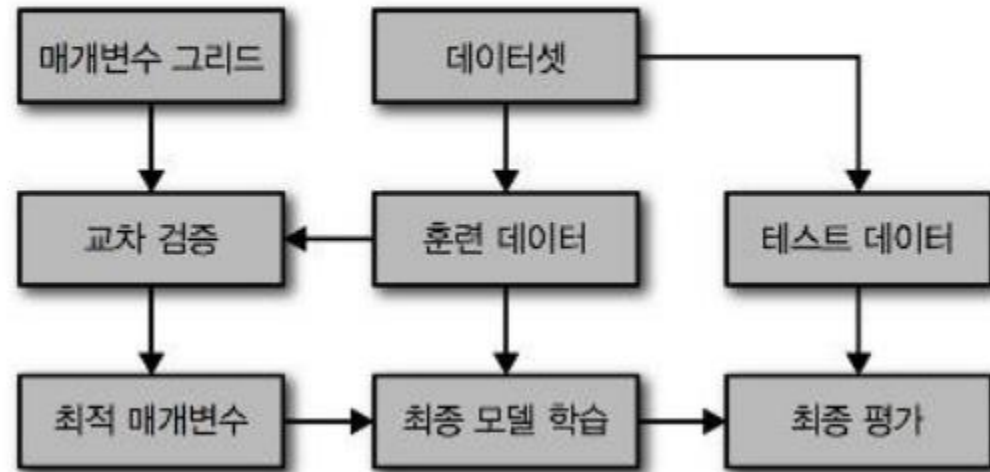


Grid Search

- 목표: 매개변수를 튜닝하여 일반화 성능을 높이는 것
- 관심 있는 매개변수들을 대상으로 가능한 모든 조합을 시도
- `Sklearn.model_selection.GridSearchCV()`
: 어떤 파라미터를 실험하고 싶은지, 어떤 값들을 시험해 보아야 하는지를 알려주면 교차검증을 사용해 파라미터 값들의 모든 가능한 조합들을 평가

모델평가 과정

매개변수 탐색의 전체 과정



실습 & 영상 😊

0.0 다중 분류 모델 성능 측정

0.1 ROC 곡선의 이해

1. Confusion matrix 실습

2. ROC, PR 곡선 실습

3. Cross Validation 실습

4. Grid Search 실습

Quest

Sklearn 의 breast cancer 데이터의 mean radius, mean texture 변수를 사용해 종양을 분류하는 모델에 대한 평가와 성능 향상

1. 두 개 이상의 분류 모델의 AUC 를 비교해 가장 좋은 모델 한 가지를 선택해 주세요 . (ROC, PR 곡선 중 데이터에 더 적합한 기준 선택)
2. 선택된 모델에 대해 5 fold 교차검증을 통한 그리드 서치를 시행해 최적의 파라미터를 도출해 주세요 .
(1 번에서 선택한 모델보다 좋은 성능을 내는 파라미터 찾기)
3. 최종적으로 최적의 파라미터로 튜닝된 모델로 score 를 산출해 주세요 !