

[머신러닝 기반 데이터 분석] 05. 모델 성능 평가하기

01. 머신러닝 수행방법 계획하기

02. 데이터 세트 분할하기

03. 지도학습 모델 적용하기

04. 자율학습 모델 적용하기

05. 모델성능 평가하기

06. 학습결과 적용하기

학습 목표

모델 성능 평가에 대해 이해한다.

모델 성능 평가에는 어떤 것들이 있는지 이해한다.

분류 모델의 모델 성능 평가에 대해 알아본다.

학습 내용

모델 성능 평가는 무엇인가?

모델 성능 평가에는 어떤 종류가 있는가?

모델 성능 평가는 왜 해야 할까?

모델 성능 평가는 어떻게 하는가?

1. 모델 성능 평가는 무엇인가?

- 모델 개발 프로세스의 필수적인 부분
- 데이터를 대표하는 최상의 모델을 찾고 미래에 **선택한 모델이 얼마나 잘 작동하는지 파악**한다.
- 과적합 모델을 만들지 않기 위해 학습에 사용되지 않은 데이터로 **모델 성능 평가** 수행

2. 모델 성능 평가에는 어떤 종류가 있는가?

Hold-out(홀드 아웃) : 큰 데이터 셋을 세개의 데이터 셋으로 나누어 수행. Training set(학습), Validation set(평가), Test set(테스트)

Cross-Validation(교차 검증) :

- 제한된 양의 데이터일 경우, 치우침이 없는 평가를 달성하기 위해 **k-fold cross-validation**(교차 검증)

3. 지도학습 - 회귀(Regression), 분류(Classification)

- 머신러닝 모델 결과가 해당 목표 값을 얼마나 잘 예측했는가?
- 일반적으로 지도학습 머신러닝 분석 모델 결과는 **모델 훈련에 사용한 훈련데이터에 편향된 결과값을 내는 경향이 많다.** 따라서 **모델 평가를 위해 평가 데이터 세트를 이용**한다.
- 회귀(수치형 데이터) 모델 평가 - **MSE, MAE, RMSE, RMLSE**
- 분류(범주형 데이터) 모델 평가 - 혼동행렬(Confusion Matrix)을 이용한 **정확도, 민감도, 정밀도**

(가) 혼동 행렬

- 혼동 행렬은 2 X 2 교차표 형태만 정의되는 것이 아니며, 3 X 3 이상의 교차표도 얼마든지 만들어 낼 수 있다.

예측 범주값

		예측 범주값	
		Y	N
실제 범주값	Y	O (TP : True Positive)	X (FN : False Negative)
	N	X (FP : False Positive)	O (TN : True Negative)

[그림 5-1] 2x2 혼동 행렬 (Confusion Matrix)

(NCS 모델 교재 참조)

(나) 분류의 주요 평가 지표

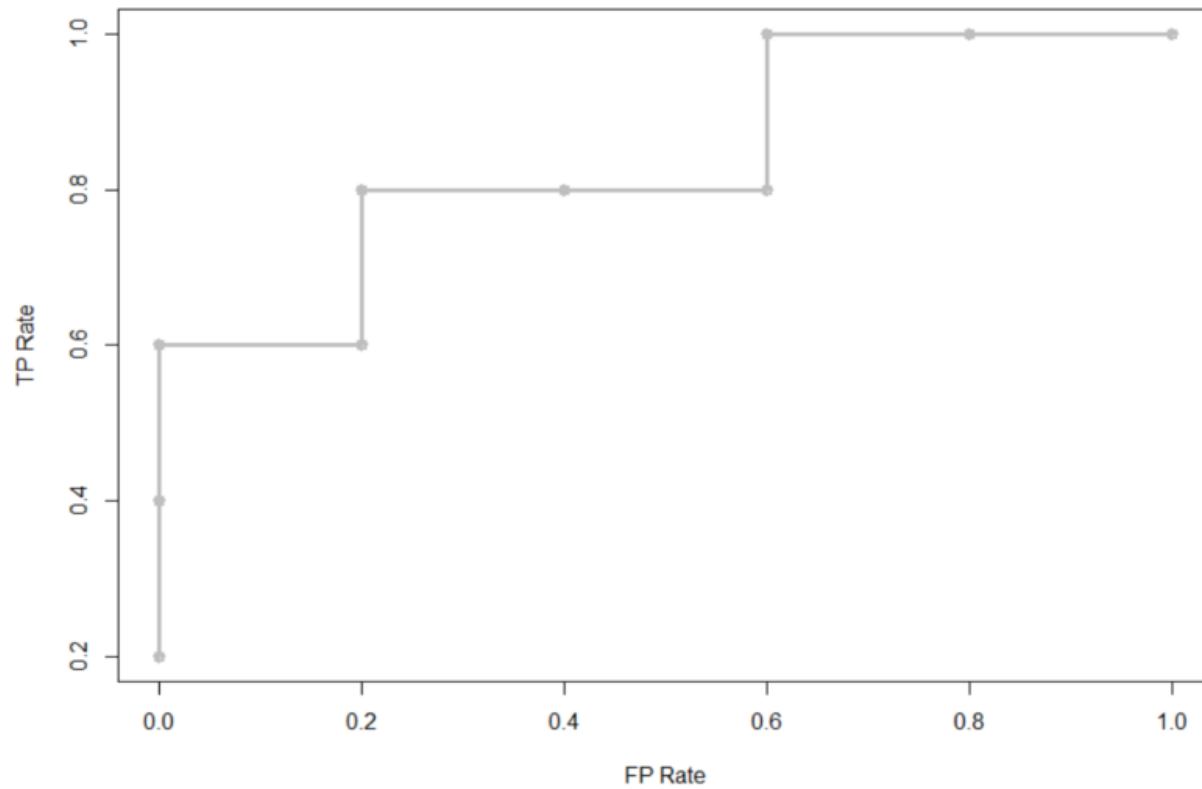
<표 5-1> 분류 목적 머신러닝 기법에서 주로 사용되는 모델 평가 지표

평가지표	계산식	지표의 의미
정확도 (accuracy)	$(TP+TN) / (TP+TN+FP+FN)$	실제 분류 범주를 정확하게 예측한 비율 (전체 예측에서 참 긍정(TP)과 참부정(TN)이 차지하는 비율)
오차비율 (error rate)	$(FP+FN) / (TP+TN+FP+FN)$	실제 분류 범주를 잘못 분류한 비율 (1-정확도 와 동일함)
민감도 (sensitivity) = 참 긍정률 (TP Rate)	$(TP) / (TP+FN)$	실제로 '긍정(positive)' 인 범주 중에서 '긍정' 으로 올바르게 예측 (True Positive)한 비율 (혹은 Recall, Hit Ratio, TP Rate 로도 지칭함)
특이도 (specificity)	$(TN) / (TN+FP)$	실제로 '부정(negative)' 인 범주 중에서 '부정' 으 로 올바르게 예측 (True Negative)한 비율

거짓 긍정률 (FP Rate)	$(FP) / (TN+FP)$	실제로 '부정(negative)' 인 범주 중에서 '긍정' 으로 잘못 예측 (False Positive)한 비율 = 1-특이도
정밀도 (precision)	$(TP) / (TP+FP)$	'긍정(positive)' 으로 예측한 비율 중에서 실제로 '긍정' (True Positive)인 비율
F-Measure (F1-Score)	$\{2 * (precision) * (recall)\} \div (precision + recall)$ $= (2 * TP) / (2 * TP + FP + FN)$	정밀도와 민감도(재현율)을 하나로 합한 성능 평가지표(정밀도와 민감도(재현율)의 조화 평균)로서, 0~1 사이의 범위를 가지며, 정밀도와 민감도 양쪽 다 클 때 F-Measure도 큰 값을 가지는 특성이 있음
카파 통계 (Kappa Statistic)	$\{Pr(a) - Pr(e)\} \div (1 - Pr(e))$	모델의 예측값과 실제값이 우연히 일치할 확률을 제외한 뒤의 값 (0~1까지의 값을 가지며 1에 가까울수록 모델의 예측값과 실제값이 정확히 일치하며, 0에 가까울수록 모델의 예측값과 실제값이 불일치하게 됨)

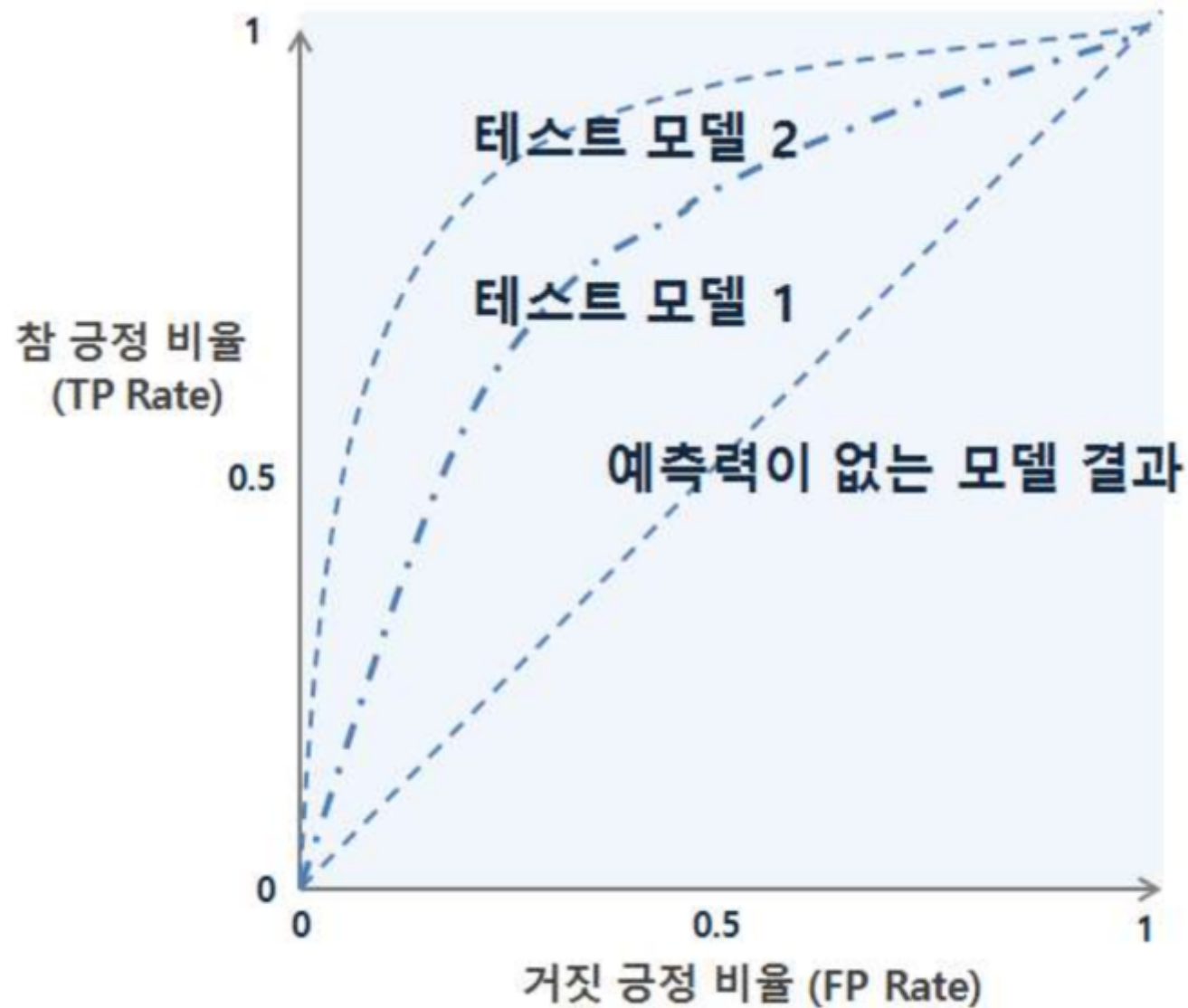
(3) ROC 곡선

ROC(Receiver Operating Characteristic) 곡선은 거짓 긍정 비율(FP Rate)와 참 긍정비율(TP Rate)간의 관계를 그래프로 나타낸 것.



[그림 5-2] 목표변수의 긍정범주 판정 기준 변화에 따른 ROC 곡선

X축을 거짓 긍정비율(FP Rate)로 하고, 그 때의 참 긍정비율(TP Rate)을 Y축으로 하여 해당 좌표를 시각적으로 표현한 것이 ROC 곡선.



[그림 5-3] 분류 머신러닝 결과의 ROC 곡선 예시

어떤 특정 분류 머신러닝 모형의 **ROC 곡선**이 좌상단에 위치할수록 모델 성능이 우수하다고 할 수 있으며, 분류 결과의 **ROC 곡선**이 대각선에 가깝게 위치할수록 참 긍정과 거짓 긍정을 제대로 구별해내지 못함을 의미.

(4) AUC(Area under the Curve)

- ROC 곡선을 이용한 평가는 직관적이기는 하다, 단, 여러모델 비교를 위해 단일 지표로 계량화가 필요하다.
- 이런 측면에서 고안된 통계치가 바로 AUC(Area under the Curve)이다.

-> ROC 곡선 아래쪽 면적에 해당되는 부분을 의미한다.

-> 값은 **0.5부터 1까지의 값**을 가진다.

-> **0.7 이상일 때, 양호한 모델로 보며, 0.8이상일 경우 뛰어난 모델로 평가된다.**