

Machine Learning Practice

Reference

- 머신러닝 딥러닝 문제해결 전략(신백균)

Machine Learning



Machine Learning

■ Classification

- Target이 범주형 데이터
- Target이 2개면 binary classification, 3개 이상이면 multiclass classification

■ Regression

- 독립변수와 종속변수 간 관계를 모델링
- 독립변수 : 영향을 미치는 변수
- 종속변수 : 영향을 받는 변수

Regression

■ Regression의 평가 지표

- MSE (Mean Squared Error) = $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- MAE (Mean absolute error) = $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- RMSE (Root Mean Squared Error) = \sqrt{MSE}
- R-squared (Coefficient of determination) = $1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$

■ 피어슨 상관계수

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

Classification

■ Confusion Matrix

True negative



False positive



False negative



True positive



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

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

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

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Classification

■ Error

■ Type 1 error

: False Positive

귀무가설이 실제로 참이지만, 귀무가설을 기각하는 오류
실제로는 negative인 것을 positive로 판정

■ Type 2 error

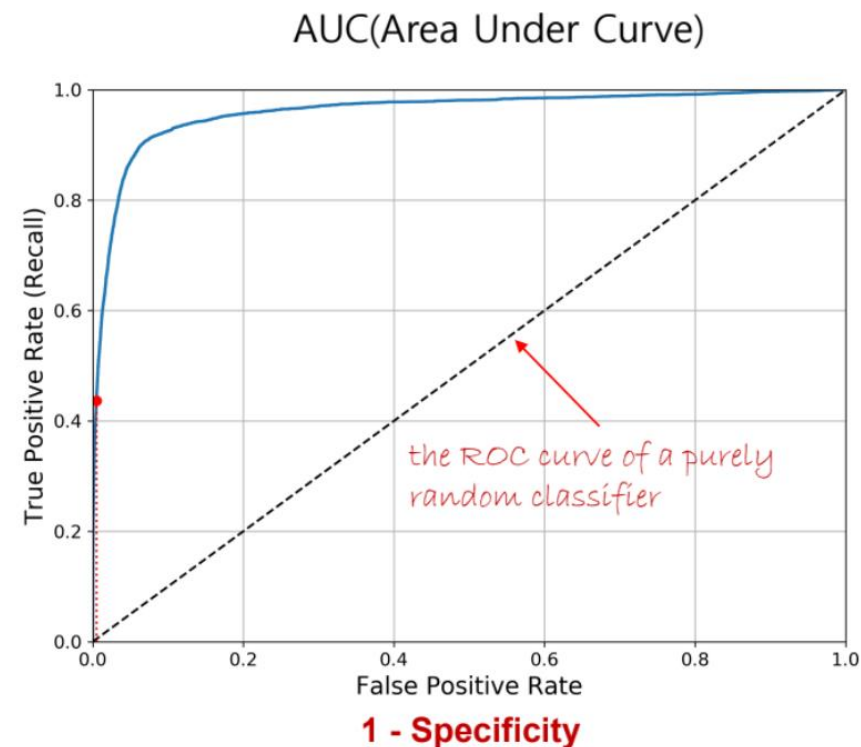
: False Negative

귀무가설이 실제로 거짓이지만, 귀무가설을 채택하는 오류
실제로는 positive인 것을 negative로 판정

Classification

■ ROC Curve

- 왼쪽 위로 갈수록 좋음
- 그래프 아래 영역의 넓이(AUC)가 클수록 좋음



Scaling

- Min-Max scaling
 - Min = 0, Max = 1이 되도록 조절
- Standardization
 - 평균이 0, 분산이 1이 되도록 조절

자전거 대여 수요 예측

■ 자전거 수요 예측(Regression)

- 워싱턴 D.C의 자전거 무인 대여 시스템 과거 기록을 기반으로 향후 자전거 대여 수요 예측
- 사용자는 한 장소에서 자전거를 대여해 원하는 만큼 타고 다른 장소에 반납
- 날짜, 시간, 요일, 계절, 날씨, 실제온도, 체감온도, 습도, 풍속, 회원 여부를 활용해 시간별 자전거 대여 수량 예측
- 데이터 분석을 위한 시각화 및 회귀, 상관관계를 나타내는 연습
- Baseline: 선형회귀와의 비교
- Ridge, Lasso, Random forest model 사용

범주형 데이터 이진분류

■ 데이터 분류 문제

- 인공 데이터 → 각 feature와 target값의 의미를 알 수 없음.
- 데이터는 모두 범주형
- 이진분류문제
- 각 feature에 대한 사전지식이 없을 때 feature에 대한 요약표를 만드는 연습
- Baseline: logistic model
- ROC AUC를 이용해 모델 최적 파라미터를 찾음

실습