

11강 모형비교평가II

통계·데이터과학과 장영재 교수



목차

01 모형비교평가 관련 R 함수

02 R 사용 예제



1. 모형비교평가가 관련 R 함수



01 모형비교평가 관련 R 함수

1 연속형 변수의 데이터 분할

■ 함수의 구조

`sample(x, size, replace=F)`

■ 기능

함수 `sample`은 내장되어 있으며 특정 크기만큼의 데이터를 임의로 선택

■ 옵션

- `x`: 표본 값의 벡터
- `size`: 선택할 표본의 수



01 모형비교평가 관련 R 함수

2 이항형 목표변수의 데이터 분할

■ 함수의 구조

`createDataPartition(y, p = 0.5, list = TRUE)`

■ 기능

함수 `createDataPartition`은 패키지 `caret`에서 제공하는 함수로서 데이터를 훈련용과 검증용 그룹으로 분할할 때 사용

■ 옵션

- `y`: 목표변수 값의 벡터
- `p`: 훈련용데이터의 비율
- `list`: list형식이면 TRUE, matrix 형식이면 FALSE



01 모형비교평가 관련 R 함수

3 예측 함수

■ 함수의 구조

`prediction(predictions, labels)`

■ 기능

함수 `prediction`은 패키지 `ROCR`에서 제공하는 함수로서 예측값을 계산할 때 사용

■ 옵션

- `predictions`: 예측 값을 담고 있는 벡터, 행렬, 또는 데이터 프레임
- `labels`: 실제 분류 수준 값(true class labels)을 담고 있는 벡터, 행렬, 또는 데이터 프레임을 입력



01 모형비교평가 관련 R 함수

4 예측 측도 함수

■ 함수의 구조

`performance(prediction.obj, measure)`

■ 기능

함수 `performance`은 패키지 `ROCR`에서 제공하는 함수로서 예측력을 계산할 때 사용

■ 옵션

- `prediction.obj`: prediction에서 발생하는 object를 입력
- `measure`: 모형평가에 사용할 측도를 지정.

"tpr"는 true positive rate의 약자로 민감도를 의미

"fpr"는 false positive rate의 약자로 1-특이도를 의미

추가로, "acc"와 "err"를 사용하여 예측정확도와 오분류율을 계산



2. R 사용 예제



다음시간 안내

12강. 군집분석 I



한국방송통신대학교