

데이터마이닝

(Data Mining)

한국방송통신대학교
정보통계학과 장영재 교수

3 강 /

회귀모형 : R을 이용한 실습

회귀모형관련 R 함수

1) 선형회귀모형 함수

➤ 함수의 구조

`lm(formula, data, na.action)`

➤ 기능

함수 `lm`은 선형모형을 적합하기 위해 사용

➤ 옵션

- `formula`: 적합할 모형의 공식으로 목표변수 ~ 입력변수의 선형결합 형식으로 구성. 입력 변수의 선형결합은 입력변수 1 + 입력변수 2 + ... 형태로 나열한다. 목표변수를 제외한 입력 데이터의 모든 변수를 입력변수로 사용할 경우 목표변수 ~ . 형태로 표현
- `data`: 사용할 데이터 프레임의 이름을 지정
- `na.action`: 결측치 처리 옵션. `na.fail`는 결측치가 있으면 오류 메시지 발생, `na.omit`와 `na.exclude`는 결측치가 있는 관측치 제외하지만 `na.exclude`는 잔차 또는 예측 값을 계산할 때 NAs를 제공

회귀모형관련 R 함수

2) 로지스틱회귀모형 함수

➤ 함수의 구조

`glm(formula, family, data, na.action)`

➤ 기능

함수 `glm`은 일반화선형모형을 적합하기 위해 사용

➤ 옵션

- `data`: 사용할 데이터 프레임의 이름을 지정
- `na.action`: 결측치 처리 옵션. `na.fail`는 결측치가 있으면 오류 메시지 발생, `na.omit`와 `na.exclude`는 결측치가 있는 관측치 제외하지만 `na.exclude`는 잔차 또는 예측 값을 계산할 때 NAs를 제공

회귀모형관련 R 함수

3) 모형 선택 함수

➤ 함수의 구조

`step(object, direction = c("both", "backward", "forward"))`

➤ 기능

함수 `step`은 AIC를 이용한 모형 선택을 위해 사용

➤ 옵션

- `object`: 함수 `lm`, `glm` 등에서 생성한 `object`를 입력
- `direction`: 변수 선택 방법 ("`both`"=단계적변수선택, "`backward`"=후진소거법, "`forward`"=전진선택법)을 지정

회귀모형관련 R 함수

4) 예측 함수

➤ 함수의 구조

`predict (object, newdata, type)`

➤ 기능

함수 `predict`은 다양한 모형 적합 결과로부터 예측 값을 계산할 때 사용

➤ 옵션

- `object`: 함수 `lm`, `glm` 등에서 생성한 `object`를 입력
- `newdata`: 예측할 변수들로 구성된 데이터 프레임을 입력
- `type`: 예측 형태를 입력하는 것으로, 목표 값을 예측할 때 `type="response"` 사용

회귀모형관련 R 함수

5) ROCR의 예측 함수

➤ 함수의 구조

`prediction(predictions, labels)`

➤ 기능

함수 `prediction`은 패키지 `ROCR`에서 제공하는 함수로서 예측 값을 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

회귀모형관련 R 함수

6) ROCR의 예측 측도 함수

➤ 함수의 구조

`performance(prediction.obj, measure)`

➤ 기능

함수 `performance`은 패키지 `ROCR`에서 제공하는 함수로서 예측력을 계산

➤ 옵션

- `prediction.obj`: `prediction`에서 발생하는 object 입력
- `measure`: 모형평가에 사용할 측도를 지정한다. “tpr”는 true positive rate의 약자로 민감도, “fpr”는 false positive rate의 약자로 1-특이도를 계산. 추가로, “acc”와 “err”를 사용하여 예측정확도와 오분류율을 계산할 수 있음

The background is a vibrant abstract composition featuring various shades of blue and purple. It includes large, soft-edged organic shapes, several circles with diagonal hatching patterns, and smaller circles with halftone dot patterns. A central white rounded rectangle contains the text.

강의를 마쳤습니다.
다음시간에는...