

2021 인공지능 소수전공

29~30차시: 로지스틱 회귀

2021.07.28 18:30~20:15

Seokhwan Yang

- 029-032_20210728_로지스틱 회귀.ipynb 참고



분석 목표에 따른 적합한 통계적 분석 기법

분석목표	설명	통계적 분석 기법
평균에 대한 검정과 추정	평균값에 대한 모델링	T검정
비율에 대한 검정과 추정	비율에 대한 모델링	직접확률계산법, F분포법
분할표의 검정	각각 2개이상의 분류 값을 지닌 2개 이상의 차원이 있고 그 결과로 하나의 측정 값이 있을 때, 분류 조합에 따라 측정값에 유의한 차이가 발생하는지를 검정	카이제곱 검정, Fisher의 직접확률 검정, 맥네마의 검정, 잔차 분석
변수들 간의 상관관계의 강도 추출	독립적으로 움직이는 두 변수들 사이의 관계(상관관계)의 강도를 상관계수로 나타내어 표시함	상관분석
변수들 간의 선형/비선형 인과관계의 형태와 강도 추출	종속적으로 움직이는 두 개 이상의 변수들 사이의 관계(인과관계)의 강도를 결정 계수로 나타내고, 각 변수의 계수를 추정해 모델화 함. 변수들은 연속적인 값일 수도 있고 분류 값일 수도 있음	회귀분석, 다중회귀분석, 로지스틱 회귀분석, 판별분석
어떤 결과에 영향을 미치는 요인들 사이의 관계와 핵심 요인의 선별	어떤 측정값에 변화 요인이 되는 값들이 세 개의 차원이라고 할 때, 각 차원들 중에 어떤 것이 측정값에 가장 큰 영향을 미치는지, 각 차원은 다른 차원의 영향력과 어느 정도 겹치는지 분석	요인분석, 주성분 분석
대상들을 여러 기준 값들에 따라 분류하고, 다차원 공간에 배치	측정값과 차원들이 있을때 차원들의 값을 기준으로 측정값들 사이의 거리를 계산해 적절하게 그룹을 짓고, 이 거리가 의미 있는 차원들로 축을 구성한 다차원 공간에 측정값들을 배치	군집분석, 다차원척도법(MDS)
차원 값들의 패턴이 비슷한 측정값과 그렇지 않은 측정값을 분류	예를 들면, 설문 항목에 대한 답변들의 패턴에 따라 비슷한 답변을 한 응답자와 그렇지 않은 응답자를 분류	대응분석
시간의 흐름에 따라 변하는 데이터를 분석할 수 있는 모델의 도출	시계열 데이터에 영향을 주는 요인을 추세요인, 계절요인, 순환요인, 불규칙요인으로 분해해서 시계열 데이터를 가장 잘 설명할 수 있는 모델을 만들고, 이 모델을 통해 미래에 대해서도 예측	시계열분석

(자료출처: 데이터 분석 전문가 가이드, 한국데이터베이스진흥원)