

I. 통계 기반 분석 모형

기술 통계

수집된 데이터를 확률·통계적으로 정리·요약하는 기초적인 통계

상관 분석

변수 간에 존재하는 상호 연관성의 정도를 측정하여 분석

회귀분석

독립변수들이 종속변수에 미치는 영향을 수정할 수 있는 통계 기법

분산 분석

집단 간 비교를 수행할 때 집단 내의 분산의 비교로 얻은 분포를 이용하여 가설검정을 수행하는 방법

주성분 분석

많은 변수의 분산 방식의 패턴을 간결하게 표현하는 주성분 변수를 원래 변수의 선형 결합으로 추출하는 통계 기법

판별 분석

집단을 구별할 수 있는 규칙이나 함수를 만들고, 다변량 기법으로 조사된 집단에 대한 정보를 활용하여 새 개체가 어떤 집단인지 탐색하는 통계기법

2. 머신러닝 기반의 데이터 분석 기법

지도학습, 비지도학습, 강화학습, 준지도학습

정답인 레이블이 포함되어 있는 학습 데이터를 통해 학습시키는 방법

유형

로지스틱 회귀

반응변수가 범주형인 경우 적용되는 회귀분석 모형

인공신경망 분석

뉴런 구조를 모방하여 만든 기계학습 모델

의사결정나무

데이터들이 가진 속성들로부터 분할 기준 속성을 판별하고, 그 속성에 따라 트리 형태로 모델링하는 분류 및 예측 모델

서포트 백터 모델

데이터를 분리하는 초평면 중에서 데이터들과 거리가 가장 먼 초평면을 선택하여 분리하는 지도 학습 기반의 이진 선형 분류 모델

랜덤 포레스트

배깅과 부스팅보다 더 많은 무작위성을 주어 약한 학습기를 생성한 후, 이를 선형 결합하여 최종 학습기를 만드는 방법

감성 분석

어떤 주제에 대한 주관적인 인상, 감정, 태도, 개인의 의견들을 텍스트로 뽑아내는 분석

4. 비지도 학습

이전 데이터에 대한 전달이 레이블이 없는 상태에서 데이터가 어떤 규

임의 데이터에 대한 정답한 데이터를 알기 위해서 어떻게

구성되었는지를 알아내는 기계 학습 기법

5. 파라미터 VS 하이퍼 파라미터

파라미터

모델 내부에서 확인이 가능한 변수로, 데이터를 통해서 산출이 가능한 값

하이퍼 파라미터

사용자가 직접 설정해주는 값

