

8장 사이파이로 공부하는 확률분포

이 장에서는 실제로 많이 사용되는 대표적인 몇 가지 확률분포의 수식과 활용을 알아본다. 우선 사이파이 파이썬 패키지를 사용하여 다양한 확률분포의 확률분포함수를 계산하고 해당 확률분포의 무작위 표본을 생성하는 방법을 알아본다.

확률분포에는 범주값을 출력하는 이산확률분포와 연속적인 값을 출력하는 연속확률분포가 있다. 여기에서는 베르누이분포, 이산분포, 카테고리분포, 다항분포 네 가지의 이산확률분포 정의와 확률분포함수를 공부하고 스팸 메일 필터링과 같은 분류문제에 어떻게 쓰이는지 살펴본다. 연속확률분포에서는 정규분포부터 공부한다. 중심극한정리는 왜 실생활에서 정규분포가 많이 사용되는지를 알려준다. 정규분포의 변형인 스튜던트 t분포, 카이제곱분포, F분포는 데이터 간의 상관관계를 따지거나 회귀분석을 하는 필수 확률분포다. 마지막으로 베타분포, 감마분포, 디리클레분포를 이용하여 베이즈 추정의 개념을 공부한다.

학습 목표

- 사이파이를 이용한 확률분포 분석 방법
- 베르누이분포, 이항분포, 카테고리분포, 다항분포
- 정규분포와 중심극한정리
- 스튜던트 t분포, 카이제곱분포, F분포
- 다변수 정규분포
- 베타분포, 감마분포, 디리클레분포