

빅데이터기초 기말과제

강의에서 사용했던 예제 코드를 활용하여 버섯의 다양한 속성들을 기반으로 버섯의 식용 여부를 판별할 수 있는 로지스틱 회귀 기반의 예측 모델을 세우고 모델의 성능을 측정하시오.

o 버섯 데이터

- 버섯 데이터는 과제와 함께 제공되는 mushroom.zip 파일을 활용하거나 UC 어바인 기계학습 저장소 페이지(<https://archive.ics.uci.edu/ml/datasets/mushroom>)를 참고하시오.
- 아래의 단계별 필수 수행 내용을 참고하여 차례로 진행하시오.

o 단계별 필수 수행 내용

1. 강의 시간에 배운 엔트로피 기반의 정보증가량(Information Gain) 방법을 이용하여 버섯의 다양한 속성 중에서 식용 가능 여부에 영향을 가장 많이 미치는 속성을 2가지 선정하시오.
2. 1의 과정에서 선정한 속성을 기반으로 식용 여부를 판단하는 로지스틱 회귀 모델을 구축하시오.
3. 훈련 및 테스트 데이터집합을 만들어 ROC 커브를 그려보고 AUC를 계산해 보시오.
4. precision-recall 커브를 그려보고 AUC를 계산해 보시오.

o 주의사항

1. 기한 내에 단계별 필수 수행 내용을 모두 포함하는 파이썬 코드를 제출하시오.
2. 가능한 단계별 필수 수행 내용을 확인하기 쉽도록 코드를 작성하시오.