**K-최근접 이웃 분류 모델 구축 및 성능 분석**

강수현

1. 데이터 불러오기

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 2개의 컬럼을 사용하여 대략적 분포 파악텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

   AI가 생성한 콘텐츠는 부정확할 수 있습니다.
2. 모델 구현텍스트, 스크린샷, 폰트, 소프트웨어이(가) 표시된 사진

   AI가 생성한 콘텐츠는 부정확할 수 있습니다.
3. Describe로 평균과 50% 구간 비교 – Weight만 큰 차이 있는 것 확인

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. Box plot으로도 확인

*텍스트, 도표, 스크린샷, 직사각형이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.*

* Weight는 Robust 스케일러, 나머지는 Standard 스케일러를 사용하기로 함

1. 8:2로 훈련 데이터와 시험 데이터 분할

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 각각 스케일 후 합치기텍스트, 스크린샷, 폰트이(가) 표시된 사진

   AI가 생성한 콘텐츠는 부정확할 수 있습니다.
2. 스케일 결과 확인

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 유클리드 / 맨해튼 적용

텍스트, 스크린샷, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 이웃수 1~30까지 비교
2. 유클리드

텍스트, 라인, 그래프, 도표이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 맨해튼

텍스트, 그래프, 스크린샷, 라인이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 최적의 이웃수를 포함한 모델 선정

텍스트, 스크린샷, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 모델 성능 분석

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

1. 결론

8:2로 데이터셋의 분할이 어떻게 되는가에 따라 결과값이 계속 바뀌었다.

40번 이상 분할과 최적모델 탐색을 반복해 보았을 때, 유클리드, 맨해튼 중 항상 더 좋았던 모델은 없었으며 최적의 이웃 수는 4이하, 정확도는 최고점이 항상 0.8 이상이 도출되었다.

가장 정밀도가 높은 물고기는 Bream이었으며 반복시행 중 전부 1.0을 기록했다. 가장 정밀도가 낮은 물고기는 Whitefish로 반복시행 중 전부 0.0을 기록했다.

Whitefish는 전체 데이터의 개수가 6으로 가장 적었고, Bream은 2번째로 많은 데이터를 가지고 있었다. 가장 개수가 많은 물고기인 Perch는 정확도가 0.7~0.9사이를 기록했는데 또한 높은 정확도를 가졌다. 충분하지 않은 데이터의 개수 때문에 이러한 결과가 나온 것이라 생각된다.