기계학습 중간고사 대체 과제

201920998 이성재

**문제정의)**

월별 기온이 총 발전량에 미치는 월별 기온효과를 분석하여 지구온난화로 인해 기온의 양극화가 심해진 계절이 발전량에도 양극화를 만드는지에 대해 확인하고자 합니다.

**데이터 선택)**

**한국가스공사\_총발전량 월별 기온효과**

총 발전량에 기온이 영향을 미친 정도를 나타내는 월별 기온효과를 나타낸 자료입니다. 2007년 1월부터 2021년 6월까지의 데이터를 포함합니다.

출처<https://www.data.go.kr/data/15066040/fileData.do?recommendDataYn=Y>

기온이 총발전량에 영향을 미친 정도가 월별로 정리되어있어서 15년 6개월간의

정보를 클러스터로 표현하기에 알맞습니다.

**최적의 클러스터 개수 찾기**

텍스트, 스크린샷, 디스플레이, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

그래프를 보았을 때 WCSS가 급격하게 감소하는 지점인 3으로 개수를 지정하겠습니다.

.

**클러스터링, 시각화**

스크린샷, 다채로움, 텍스트이(가) 표시된 사진

자동 생성된 설명

**결과해석)**

3개의 군집으로 나눈 결과입니다.

군집을 보았을 때 발전량에 영향을 미친 값인 temp(기온효과)가 0.10 근처 값의 분포, 밀집이 줄어들고 0.10을 기준으로 잡았을 때 위 아래로 양극화되는 것이 보입니다.

또한 기온효과가 큰쪽으로 밀집되고 있는 것이 보입니다.

지구온난화로 인해 겨울을 더욱 추워지고 여름은 더욱 더워지는 이상기후로 인해 계절에 따른

가스발전량의 양극화가 나타나는것으로 보입니다.

지구온난화로 인한 기후변화가 발전량에 영향을 준 것을 확인할 수 있었습니다.

**클러스터 중심점 분석)**

**텍스트, 전자제품, 스크린샷, 소프트웨어이(가) 표시된 사진

자동 생성된 설명**

중심점 분석을 위해

centroids = kmeans.cluster\_centers\_

와

클러스터 시각화에

plt.scatter(kmeans.cluster\_centers\_[:, 0], kmeans.cluster\_centers\_[:, 1], s=200, c='red', marker='X', label='Centroids')

위의 코드를 추가 했으나 적은 데이터 수와 밀집되지않은 데이터 탓에 이러한 결과가 나온것같습니다.

**실루엣 점수)**



여기서 실루엣 점수가 0.03으로 매우 낮게 나왔으므로 클러스터링 결과가 중간 수준으로 평가됩니다. 이는 기온은 서서히 변화하는 특성상 기온효과 또한 수년간의 자료를 사용하여도

컴퓨터의 입장에선 아주 커다란 변화가 볼 수 없으므로

클러스터 간의 분리가 뚜렷하지 않은 것이 이유로 보입니다.

**감사합니다.**