

📌 발표자료는 ppt 형식으로 만든 후에, 최종적으로 pdf로 변환해서 전달해 주세요.

📌 발표는 30분 ~ 35분 분량으로 준비해주시면 됩니다. (대략 45 슬라이드 이상)

📌 템플릿은 카페에 있는 공용 템플릿을 활용해 주시길 바랍니다.

-> <https://cafe.naver.com/ewhaeuron/321>

📌 발표자료는 세션 전날 (11/17) 23:59까지 운영1 !조승연!님께 제출해 주세요.

📌 11주차 복습과제를 출제해 주셔야 합니다.

**[복습과제 출제 가이드]**

- 문항 수) 20문제 정도

- 각 파트 당 3~4문제 정도씩 출제해 주시면 됩니다.

- step을 잘게 나누셔도 상관은 없습니다만, 2시간 안쪽으로 과제를 끝내실 수 있을 정도의 분량으로만 출제 해주세요.

- 데이터셋 제공처

↳ <https://learning-sarah.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D%EB%94%A5%EB%9F%AC%EB%8B%9D-%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%85%8B-%EC%A0%9C%EA%B3%B5-%EC%82%AC%EC%9D%B4%ED%8A%B8>

- 문항 구성)

- 코테와 비슷한 형식으로 출제해 주시면 됩니다. 또는 아래 7기 자료를 참고하셔도 됩니다.

[https://github.com/Ewha-Euron/7th-ML/tree/Week\\_13](https://github.com/Ewha-Euron/7th-ML/tree/Week_13)

(형식만 참고하시고, 문제는 새로 출제해 주셔야 합니다.)

- 교재에 있는 내용을 참고하여 약간만 응용해 주시면 됩니다.

(같은 코드/로직인데 데이터셋만 바꾼다던가, 또는 활용하는 모델을 변경한다던가 등)

- 각자 세션 때 발표하는 파트에 대해서 출제해 주시면 좋을 듯 합니다.

- 문제지와 답안지를 모두 제작해 주셔야 합니다. (각각 별도의 ipynb 파일로)

- 문제 조건의 경우 마크다운 셀로 작성하는 것을 추천(또는 code 셀의 주석)

- 답안지의 경우 과제 마감 후 부원분들께도 배포될 자료이니, 설명은 최대한 자세히 적어주시면 감사하겠습니다.

- 출제하시고 ipynb 합치실 때 문제 번호 형식도 맞춰주시는 게 좋습니다.

- 출제 이후 운영진에서 검수 후, 수정이 필요한 부분이 있다면 연락 드리겠습니다.

📌 발표팀은 11주차 예/복습과제를 제출하지 않으셔도 됩니다.

아래 정리된 주요 내용을 중심으로 발표를 진행해주시면 됩니다!

## <7장. 군집화>

### (A)

#### 7-0. 군집화 개념

- 발표 시작 전에 군집화와 관련하여 알아야 할 부분들을 정리해 주세요.
- 간단하게 키워드 위주로 정리해 주시면 됩니다~
  - 지도학습 vs 비지도학습
    - ↳ <https://zhining.tistory.com/25> (강화학습은 다루지 말아주세요.)
  - 군집화의 정의
  - 군집화의 종류: 비계층적 군집화(>> 거리 기반 군집화 vs 밀도 기반 군집화) vs 계층적 군집화
    - ↳ <https://sosomeasy.tistory.com/214>
  - 하드 군집 vs 소프트 군집
    - ↳ <https://sosomeasy.tistory.com/243>

#### 7-1. KMeans

- ↳ ! 개념정리 !
- 대표적인 거리 기반 군집화 알고리즘입니다.
- KMeans 알고리즘 동작 방식을 자세히 설명해 주세요.
- 주요 파라미터와 그 역할을 정리해 주세요.
  - (n\_clusters, init, max\_iter, cluster\_centers)
- 군집화 알고리즘을 테스트를 위한 데이터 생성
  - ↳ make\_blobs(), make\_classification()
  - 주요 파라미터 정리 해주세요~

#### 7-2. 군집 평가

- 실루엣 분석
- 좋은 군집화란 무엇인지

#### + 추가 내용

- 이너셔(inertia), Elbow Method
  - ↳ <https://humankind.tistory.com/22>

### (B)

#### 7.3 평균 이동

- KMeans 군집화와 비슷한 거리 기반 군집화 알고리즘입니다.
- 알고리즘 동작 방식을 자세히 설명해 주세요.

#### 7.4 GMM

- 알고리즘 동작 방식을 자세히 설명해 주세요.
- GMM과 KMeans의 차이점

#### 7-5. DBSCAN

- 대표적인 밀도 기반 군집화 방법입니다.
- 주요 파라미터 & 데이터 포인트
- DBSCAN 적용 예제

(C)

+ 추가 내용

- 병합 군집, 계층적 군집
  - ↳ 병합 군집 개념: <https://kolikim.tistory.com/31>
  - ↳ linkage 옵션: <https://bizzengine.tistory.com/152>

7-6. 군집화 실습 - 고객 세그먼테이션

- 군집화가 가장 많이 활용되는 분야(task)로, 마케팅 분야 쪽에서 특히 주목하고 있는 과제입니다.
- 고객 세그먼테이션의 기본적인 기법은 RFM 기법입니다.
  - 따라서, 군집분석을 수행하기 전에 RFM에 맞게 데이터를 가공하는 과정이 필수적입니다.
  - 교재에서 데이터 가공을 어떻게 하고 있는지에 주목해서 발표 구성해 주시면 좋을 듯 합니다.