|  |  |
| --- | --- |
| 교육 제목 | 데이터와 알고리즘 |
| 교육 일시 | 2021.11.23 |
| 교육 장소 | 영우글로벌러닝 |
| **교육 내용** | |
| 오전 | 소프트벡터의 점선으로부터 경계까지의 거리가 마진이다  훈련세트에서 모델을 훈련하면 검증 세트로 모델을 평가한다  군집 - 비슷한 샘플끼리 그릇으로 모은다  클러스터 - 군집 알고리즘에서 만든 그룹  평균 알고리즘 작동방식  1. 무작위로 k개의 클러스터 중심 잡기  2. 샘플에서 가까운 클러스터 중심을 찾아 그 클러스터를 샘플로 지정  3. 샘플의 평균값으로 중심변경  4. 클러스터 중심의 변화가 없을때까지 2번으로 돌아가 반복 |
| 오후 | 컴퓨터가 잘 이해하는 구조로 되어있는 데이터를 정형 데이터  이와 반대되는 데이터를 비정형 데이터라고 한다  비정형 데이터는 머신러닝으로 모델을 만들기가 어렵다  정형데이터를 다루는데 뛰어난 성과의 알고리즘은  결정트리를 기반으로 한 앙상블 학습이다  부트스트랩 샘플은 중복샘플을 뽑을 수 있다  엑스트라 트리는 100개의 결정트리를 훈련한다  훈련세트를 사용한다는데 랜덤포레스트와는 차이가 있다  그레이디언트 부스팅은 깊이가 얕은 결정트리로 이전트리의 오차를 보완하는 방식이다  히스토그램 기반의 그레이디언트는 머신러닝 알고리즘 중 가장 인기가 높고 안정적이다 |