|  |  |
| --- | --- |
| 날짜 | 2020년 3월 6일 |
| **1. 이 번주 연구 관련 학습 내용 정리**   * **사이킷런을 이용한 머신러닝 분류 모델(chapter 3)** * **강의 chapter1,2 수강**   <https://www.udemy.com/course/complete-machine-learning-and-data-science-zero-to-mastery/> | |
| **2. 다음 주 학습 계획**   * **듣던 강의 chapter 3 수강** * **머신러닝 교과서 책 chpater4-3까지 공부하기** | |
| **3. 이번주 학습 내용 정리**  **사이킷런과 붓꽃데이터셋을 이용하여 퍼셉트론을 훈련시켰다. 적절한 학습률을 찾기 위해 에포크마다 훈련세트를 섞어서 사용했다. 퍼셉트론 알고리즘을 이용할 경우 선형적으로 구분가능한 데이터셋은 구분이 가능하지만 그렇지 않은 경우는 불가능하여 이 알고리즘을 실전에서 자주 사용하진 않는다고 한다.**  **그래서 사용하는 방법으로는 로지스틱 회귀 알고리즘이 있다. 이는 시그모이드 함수를 이용하여 선형적이지 않은 (1과 0으로 나눌 수 없는) 값들도 분류가 가능하다. 로지스틱과 퍼셉트론 방법을 두개의 데이터셋으로 나눠지는 데이터를 훈련시켰을 때는 차이가 거의 없다.**  **과대적합이란 훈련데이터로는 잘 동작하지만 새로운 데이터로는 잘 일반화가 되지 않는 경우로 이 경우에는 규제라는 방법을 사용하여 과대적합을 방지한다.**  **서포트 벡터 머신은 최적화 대상의 마진을 최대화하는 방법으로 마진이란 클래스를 구분하는 결정경계와 그 경계에 가장 가까운 훈련 샘플 사이의 거리이다. 이를 사이킷런을 이용하여 구현하였고 비선형 문제도 해결할 수 있었다.**  **강의에서는 튜토리얼과 비슷한 기본 강의 였다. 주요 내용으로는 머신러닝이랑 무엇이며 이 강의를 통해 배울게 될 내용에 대해 설명하였다. 이 강의를 보고 내가 내린 머신러닝의 정의는 기계가 데이터를 정보로 받아들여 학습을 하고 그에 따른 결과를 만들어내는 것이라고 생각한다.** | |