연구 내용 요약 및 계획

|  |  |
| --- | --- |
| 날짜 | 2020년 2월 28일a |
| **1. 이 번주 연구 관련 학습 내용 정리**   * **머신러닝 교과서 책 읽고 따라하기** | |
| **2. 다음 주 학습 계획**   * **읽던 책 마져 읽고 실습 진행하기** | |

전처리 : 원본 데이터에서 학습하기 쉽도록 하기 위해 특징을 추출해내서 새로운 데이터셋을 만드는것

예를 들면 차원 축소 방법이 있다. 이는 저장공간도 덜 필요 할 뿐만 아니라 학습 알고리즘의 속도도 증가한다. 그러하면 예측 성능도 높아질 수 있다.

데이터 셋과 관련없는 특성이 매우 많을 경우 = 신호 대 잡음비가 낮을 경우

하이퍼파라미터 : 데이터에서 학습하는 파라미터가 아니라 모댈 성능을 향상하기 위해 사용하는 다이얼.

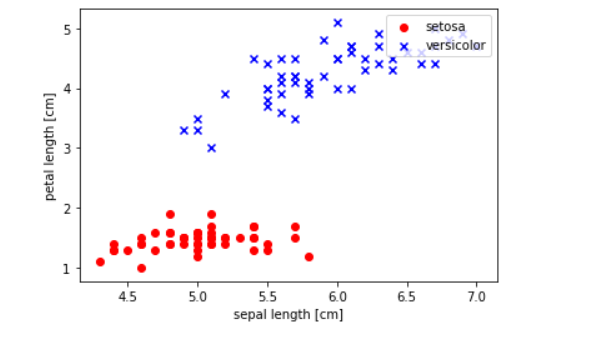
* 아나콘다를 이용한 tensorflow, python환경 구축 완료

Data science made three parts. Coding, statics, domain(business)

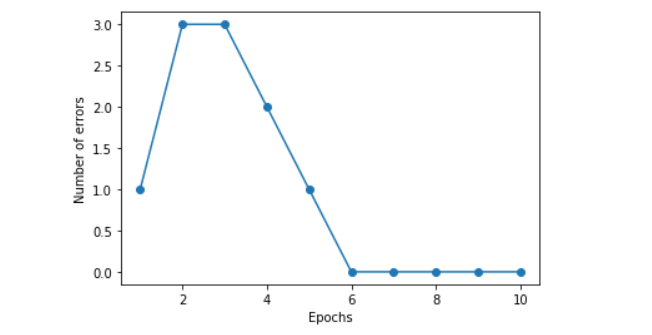
DS and statcis use data. But different background.

퍼셉트론이 정확히 예측 : 0, 잘못 분류하면 : 양수나 음수

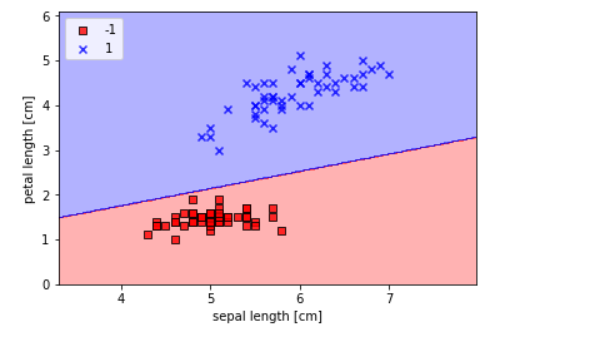
잘못 분류했다면 다음번에는 올바르게 분류하기 위해 결정경계가 더 크게 움직인다. 즉, 가중치 업데이트는 이전의 x값에 비례한다.



봇꽃 데이터 셋을 가져와서 setosa, versicolor의 꽃받침 길이와 꽃잎길이를 50개를 iloc을 이용하여 추출하여 그 값으로 scatter를 이용하여 산점도를 그린 결과.



그린 결과를 perceptron에 훈련시켜 경곗값을 구하도록 한다. 그 때 에러가 발생하는 회수를 출력. 6번째 데이터를 학습할 때부터 결과에 수렴한다.



학습한 결과를 토대로 경곗값을 출력한 결과. 데이터가 둘로 완벽하게 나눠지므로 성공한 결과이다.