

2017-06-24 (1주차-2)

1. numpy를 이용한 선형대수
<https://datascienceschool.net/view-notebook/cd600e9627cd4f16bc4f6078231ab28c/>
2. 벡터와 행렬의 연산
<https://datascienceschool.net/view-notebook/3f44cfdda2874080a9aa6b034c71d5ec/>
3. 행렬의 성질
<https://datascienceschool.net/view-notebook/d6205659aff0413797c22552947aec83/>
4. 연립방정식과 역행렬
<https://datascienceschool.net/view-notebook/927d7f7972dd434ead9d294169ae7f34/>
5. 선형대수와 해석기하
<https://datascienceschool.net/view-notebook/dd1680bfbaab414a8d54dc978c6e883a/>
6. 질문
 - text 를 vector화 하게 되면 너무 큰 sparse vector 다른 feature와 concat하는 좋은 방법은?
-> 현재 TFIDF를 사용하고 있음으로 값의 범위는 0~1이고 따라서 concat할 변수의 scaling을 0~1사이로 하는 것은 일반적으로 맞다.
-> 가중치를 주기 위해서 scaling 을 변화시키는 것은 옳바르지 않다. weight 로 가중치를 주는 것이 바람직함
 - another class 생성방법
 - sklearn svm에서 ovo option이 존재 one vs one classifier에 svm을 인자로 넘기는 것은 중복이 아닌가?
 - $y = Xw$ (X : batch size), nn에서 $y = Wx$ 그렇다면 nn에서 전체 행렬식표현은?
 - 텍스트 피처의 경우 데이터보다 feature가 많을 수 있지 않나? 최소자승법으로 접근 X?
 - corpus 변화
-> 질문정리