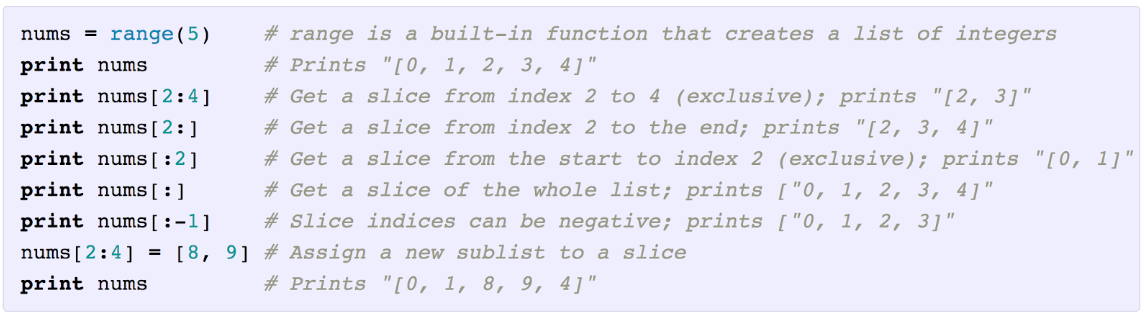
ML lab 04-2 : TensorFlow로 파일에서 데이터 읽어오기

파일이 커짐으로써 data를 로딩해올 필요가 있음.

‘,’로 나누어진 csv파일을 주로 사용함. 파일 안에 데이터들이 같은 데이터 타입이어야 함.

slicing





file이 굉장히 커서 메모리에 한번에 올릴 수 없을 때 queue runner를 이용함. 여러 개의 파일에서 읽어와서 queue에 쌓아놓은 후 batch만큼씩 reader로 읽어온다. 굉장히 큰 데이터를 사용할 때 유용함.

1. filename을 queue에 list해준다.
2. reader를 정해서 읽어준다.
3. value를 어떻게 parsing할 것인가 decode\_csv를 사용.

batch를 만들어서 읽어온다. 펌프처럼 데이터를 읽어옴. batch\_size : 한번에 몇 개의 데이터를 읽어올 것인가?



shuffle\_batch를 이용하면 batch의 순서를 섞을 수 있음.