

# 딥러닝 실습 (Word Embedding)

“ 실습

Visualization

INDEX



## 실습 개요

- ❑ **Step 1)** Preprocessing되어있는 Wikipedia Korean corpus를 대상으로 word2vec 학습을 진행함  
(특수문자 및 기호 제거, 숫자 -> '0'으로 )
- ❑ **Step 2)** 학습한 모델을 이용하여 visualization



## Word2vec 학습하기

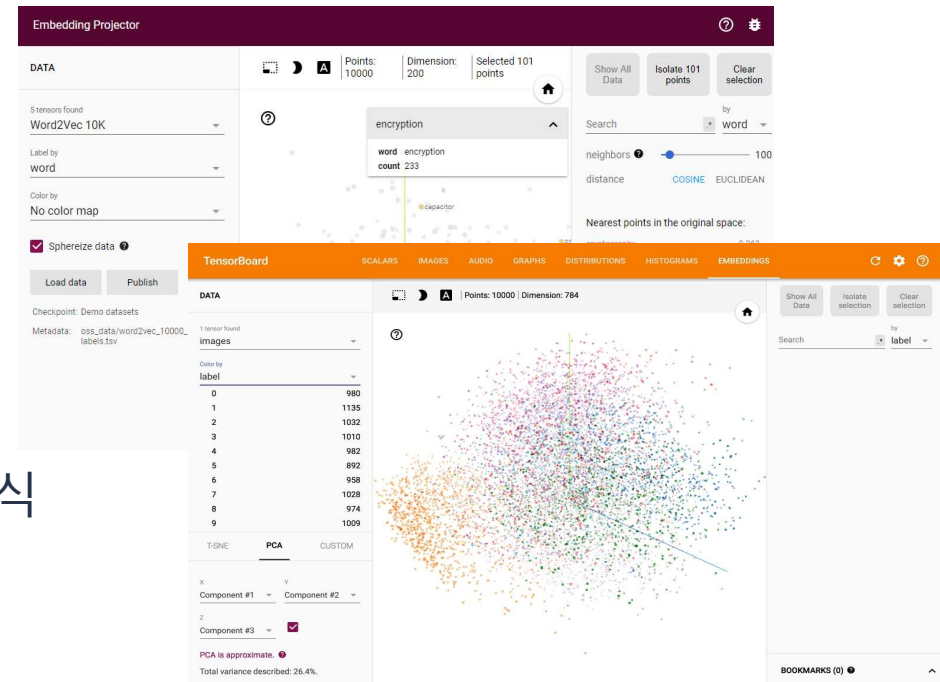
- ❑ Gensim word2vec library 사용 (CPU 버전)
  - ❑ 심플한 사용이 강점
- ❑ 그 외에 word2vec 참고 자료
  - ❑ Tensorflow word2vec (GPU 버전 사용가능)
    - ❑ <https://www.tensorflow.org/tutorials/representation/word2vec>
  - ❑ Word2vec numpy 구현
    - ❑ <https://towardsdatascience.com/an-implementation-guide-to-word2vec-using-numpy-and-google-sheets-13445eebd281>

# “ Visualization



# Tensorboard

- ❑ Google에서 제공해주는 visualization tool
- ❑ 온라인상에서 원하는 파일을 업로드하여 시각화하는 방식
  - ❑ Embedding projector (url로 접속)
- ❑ 로컬 pc내에서 시각화하는 방식
  - ❑ Tensorboard (설치 후 이용가능)





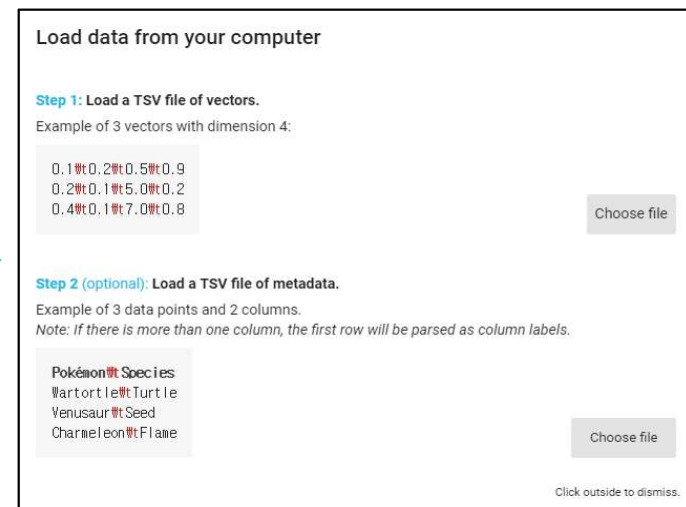
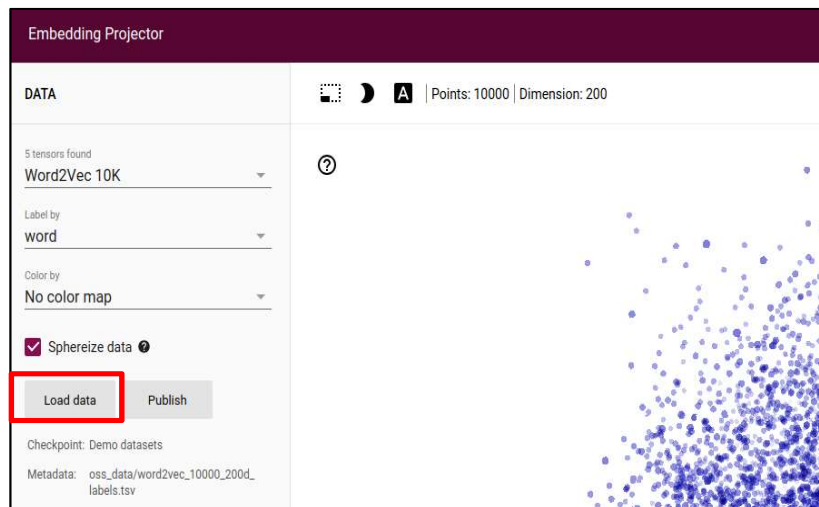
# Tensorboard

- ❑ 코드 실행을 위해 필요한 파일
  - ❑ wiki\_korean\_model
- ❑ 시각화를 위해 해당 코드에서 도출되는 output
  - ❑ Wordvectors (단어 벡터)
  - ❑ Labels (vocabulary)



# Tensorboard

- ❑ Tensorboard 에서 시각화 확인을 위해 필요한 파일
  - ❑ Word vectors file (.tsv)
  - ❑ Metadata file (.tsv)







## Tensorboard (Embedding projector) 접속

□ <http://projector.tensorflow.org/>

The background features a large, light blue trapezoidal shape on the left side. Below it is a dark blue trapezoidal shape. In the bottom right corner, there is a small orange trapezoidal shape. The text "Thank you" is centered within the dark blue shape.

**Thank you**