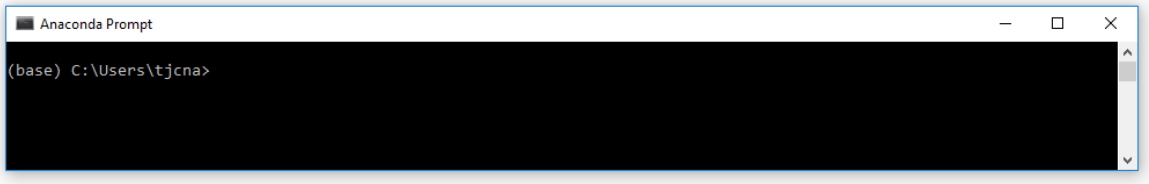
Nanyang Technological University

IS6751 Text and Web Mining
Lab: The CBOW Classifier Model & Word Embeddings with gensim

In this lab, you will learn about the CBOW (Continuous BOW) Classifier Model and Word Embeddings with gensim.

If necessary, upgrade gensim to use the Word2Vec library.

Start | Anaconda3 | Anaconda Prompt



Type the following command

>>> pip install gensim -U

1. Running the Jupyter Notebook

1.1. Copy the **notebook files** downloaded from the class website to **C:\Users*youraccountname*\documents\python\IS6751**

1.2 Open a Jupyter Notebook: Data Preparation for the CBOW Classifier Model

- File – Open …, **5_2_munging_frankenstein-v2.ipynb** in **C:\Users***youraccountname* **\documents\python\IS6751\wk7-cbow-chapter_5\5_2_CBOW**
- Run the notebook document step-by-step (one cell a time)

1.3 Open a Jupyter Notebook: the CBOW Classifier Model

- File – Open …, **5_2_Continuous_Bag_of_Words_CBOW-v2.ipynb** in **C:\Users***youraccountname* **\documents\python\IS6751\wk7-cbow-chapter_5\5_2_CBOW**
- Run the notebook document step-by-step (one cell a time)

1.4 Open a Jupyter Notebook: Word Embeddings with gensim

- Download **GoogleNews-vectors-negative300.bin.gz** (*https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit?usp=sharing*), and unzip and save it in **C:\Users***youraccountname* **\documents\python\IS6751\wk7-cbow-chapter_5\gensim**
- File – Open …, **gensim_word_embedding_v4.ipynb** in **C:\Users***youraccountname* **\documents\python\IS6751\wk7-cbow-chapter_5\gensim**
- Run the notebook document step-by-step (one cell a time)