

## BDC4CM Reading materials and software libraries

### Reading materials

- Learn PPMI dataset
  - <http://www.ppmi-info.org/>
  - The Parkinson Progression Marker Initiative (PPMI). Progress in neurobiology, 2011.
- Learn Parkinson's Disease prediction
  - Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations.
- Learn cognitive impairment prediction
  - Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study. The Lancet Neurology, 2017.
- Learn about Linear Classification
  - LIBLINEAR: A Library for Large Linear Classification, Journal of Machine Learning Research, 2008
- Learn about 1-based feature selection
  - Compressive Sensing, IEEE Signal Processing Magazine, 2007
  - [http://scikit-learn.org/stable/modules/feature\\_selection.html](http://scikit-learn.org/stable/modules/feature_selection.html)
- Learn about MeSH
  - [https://en.wikipedia.org/wiki/Medical\\_Subject\\_Headings](https://en.wikipedia.org/wiki/Medical_Subject_Headings)
  - NLM MeSH <https://www.nlm.nih.gov/mesh/>
- Learn about JSON Lines format
  - <http://jsonlines.org/>
- Learn about Source normalized impact per paper (SNIP)
  - <https://blog.scopus.com/posts/journal-metrics-in-scopus-source-normalized-impact-per-paper-snip>
- Learn about downloading bibliographic records from PubMed
  - NCBI E-utilities <https://www.ncbi.nlm.nih.gov/books/NBK25497/>
  - Biopython Entrez <http://biopython.org/DIST/docs/api/Bio.Entrez-module.html>
- Learn about Word2vec

- <https://en.wikipedia.org/wiki/Word2vec>
- Genism word2vec module  
<https://radimrehurek.com/gensim/models/word2vec.html>
- Learn about NLTK tokenization
  - <http://www.nltk.org/api/nltk.tokenize.html>
- Learn about tf-idf and LDA
  - <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>
  - [https://en.wikipedia.org/wiki/Latent\\_Dirichlet\\_allocation](https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation)
  - Gensim tf-idf <https://radimrehurek.com/gensim/models/tfidfmodel.html>
  - Gensim LDA <https://radimrehurek.com/gensim/models/ldamodel.html>
  - Gensim similarity queries <https://radimrehurek.com/gensim/tut3.html>
- Learn about sklearn TSNE
  - <http://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

## Python libraries

- Liblinear
  - Download: <https://www.csie.ntu.edu.tw/~cjlin/liblinear/>
  - Installation:  
<https://github.com/cjlin1/liblinear>  
or using Sci-kit learn directly

### MacOS

- ✓ Press `Command+Space` and type `Terminal` and press `enter/return` key.
- ✓ Run in Terminal app:  
`ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)" < /dev/null 2> /dev/null`  
and press `enter/return` key. Wait for the command to finish.
- ✓ Run:  
`brew install liblinear`
- ✓ Done! You can now use `liblinear`.

### Ubuntu 14.04

- ✓ `sudo apt-get update`
- ✓ `sudo apt-get install python-Liblinear`

Then copy the .py files `liblinear.py` and `liblinearutil.py` into project folder. It can be imported by:

```
from liblinear import *
from liblinearutil import *
```

- Natural Language Toolkit (NLTK)

- Installation instructions(package) <http://www.nltk.org/install.html>
- Installation instructions(data) <http://www.nltk.org/data.html>
  
- Biopython
  - Installation instructions <http://biopython.org/wiki/Download>
  
- gensim <https://radimrehurek.com/gensim/>
  - Installation instructions <https://radimrehurek.com/gensim/install.html>
  
- Scipy, Numpy, Pandas, Matplotlib, Sci-kit learn
  - <https://www.scipy.org/>
  - <http://www.numpy.org/>
  - <http://pandas.pydata.org/>
  - <https://matplotlib.org/>
  - <http://scikit-learn.org/>