Week 6
MA5851 – Data Science Master Class 1
[Model Development and Deployment]

Dr Mostafa Shaikh

mostafa.shaikh@jcu.edu.au

online.jcu.edu.au

JAMES COOK UNIVERSITY
AUSTRALIA

Cairns
Singapore
Townsville

Topic

- Notebook export as PDF
- Code Repository – github, SourceTree
- Deploy on cloud – Heroku, AWS, Azure
- Local ML model development
  - PyCaret
  - Kaggle (https://www.kaggle.com/mirichoi0218/insurance/data)
  - Flask (web development)
- Managing the code through git – version control
- Test local app
  - As python notebook
  - As local web app: frontend, backend, testing on localhost
- Deployment steps: readme, license, **requirements.txt, Procfile**
- Deploy on Heroku from github repo
- Testing the application

Notebook to PDF

- On Mac: https://www.tug.org/mactex/mactex-download.html

- pip install -U notebook-as-pdf

Additional setup for Chromium

- pyppeteer-install

command prompt to convert the notebook

- jupyter-nbconvert --to PDF <my_notebook.ipynb>
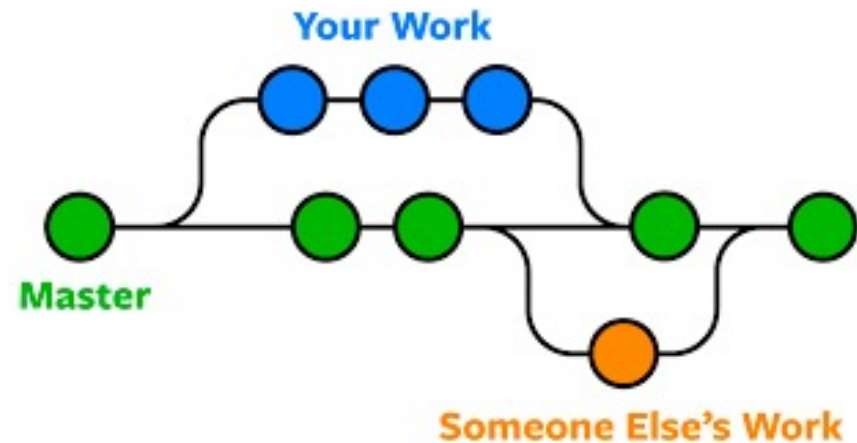
You can also do "Download as" PDF from the Notebook

Version Control

- https://github.com/
- https://bitbucket.org/
- https://www.sourcetreeapp.com/

Overview

- Creating a local repository
- Adding files
- Commiting changes
- Viewing history
- Viewing a diff
- Working copy, staging, and repository
- Deleting files
- Cleaning the working copy
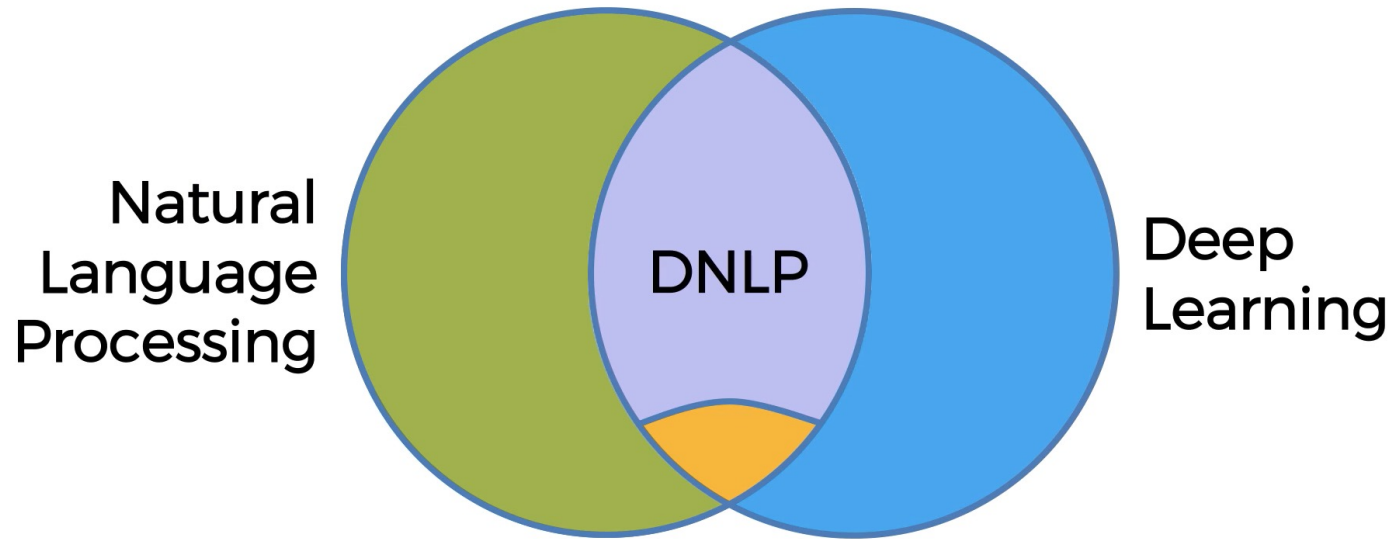- Ignoring files with .gitignore

Important Git Commands

✓ git log

✓ git log --graph -oneline

✓ git status

✓ git add

✓ git branch

✓ git checkout (branch or file -r)

✓ git diff

✓ git reset

✓ git tag (stable version)
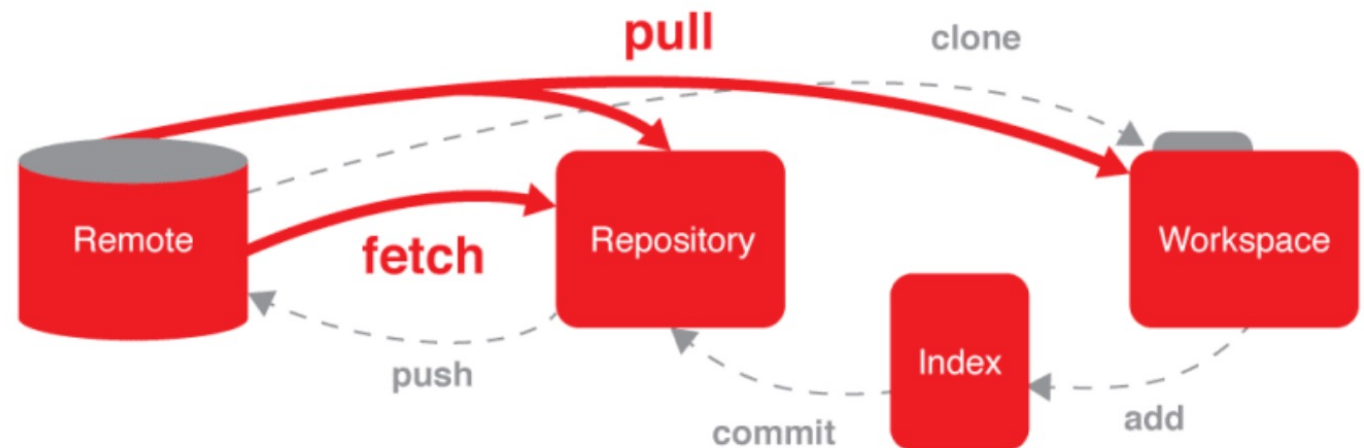
Classic vs Deep Learning Model

- If-else rules (eg, Q&A, Chatbot)
- Speech Recognition based on audio Frequency component analysis
- Bag of words model (for classification)
- CNN for text classification
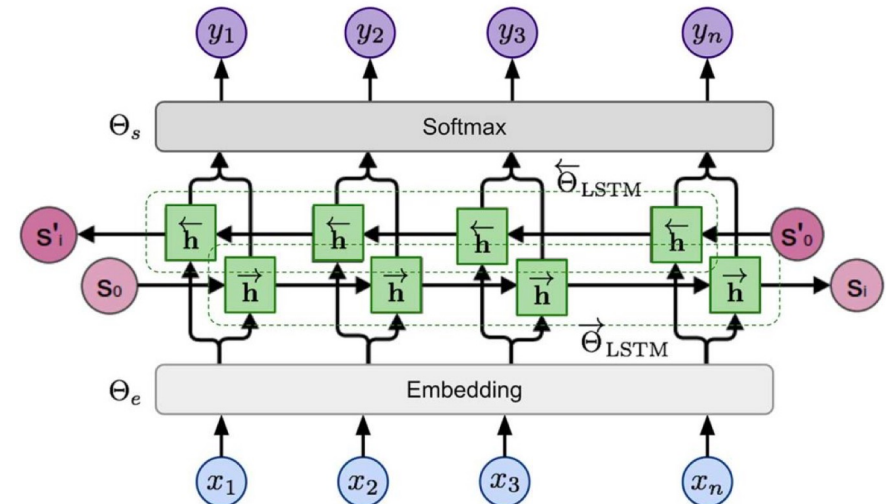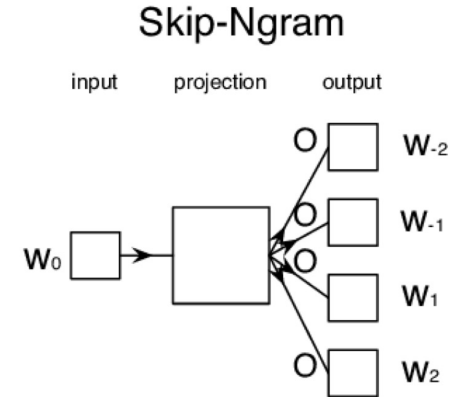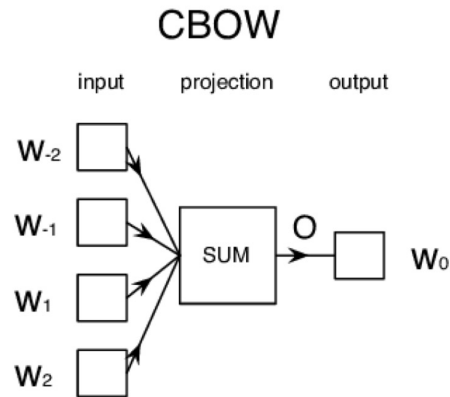- Seq2Seq model (eg, Generative Pre-trained Transformer 3, GPT-3

Embeddings

Word Embedding

- Word2Vec
  - CBOW
  - Skip-Ngram

- GloVe: https://nlp.stanford.edu/projects/glove
- FastText

Sentence Embedding
- ELMo: Embedding from Language Model

- BERT: Bidirectional Encoder Representations from Transformers

Pre-trained Language Models

- spaCy (word2vec)
- fastText
- GloVe
- Poincare
- Numberbatch
- flair