Workshop 1

BUSA90542 Machine Learning

Tutorial structure

- Review concepts from the morning lecture
  - Class discussion
  - Pen/paper questions
  - Coding questions in Jupyter Notebook

Please ask questions at any time!

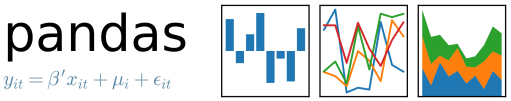
Today's agenda

1. Intro to the Python ecosystem for ML
2. Intro to Jupyter Notebook
3. Worksheet
    1. Evaluation measures
    2. Cross-validation
    3. Naïve Bayes classifiers

Python ML ecosystem

The SciPy stack

- Library for working with large multidimensional arrays
- High-level functions for arrays



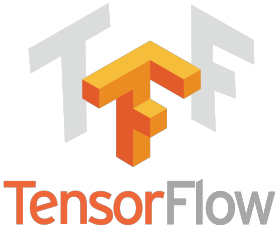
- Library for analysis and manipulation of tabular data
- Provides similar functionality to DataFrames and dplyr in R

- Machine learning library
- Includes implementations of most models covered in this course (exception: neural nets)

- Scientific computing library
- Functionality includes: statistics/random number generation, linear algebra, optimisation, special functions

- 2D plotting library
- Provides similar interface to MATLAB

Other libraries in the ecosystem



- Library for numerical computations using data flow graphs
- Often used for neural nets
- Supports GPU, TPU acceleration

- Python library built on top of C++ backend of computational library Torch.
- Similar to TensorFlow, used for NN and supports GPU, TPU acceleration

- High-level neural net library written in Python
- Supports various backends: TensorFlow, CNTK and Theano

- A probabilistic programming language written in C++
- Great for smaller-scale statistical modelling
- Has a Python API

- Cluster computing framework
- Supports scalable machine learning through Spark MLlib
- Has a Python API

- A probabilistic programming language written in Python and built on top of Theano
- More “Pythonic” than Stan

Installing Python on your device

- We recommend the Anaconda Python distribution
- Specialises in scientific computing
- Cross-platform: available for Windows, Mac, Linux
- anaconda.com/distribution
-

Jupyter Notebook

- Interactive notebooks
  - Text (in Markdown)
  - Code blocks
  - Visualizations
- Implemented as a web app
- Supports many languages (including Python)
- jupyter.org
- Included in Anaconda

Intro to Jupyter Notebook

Worksheet