MACHINE LEARNING USING JUPYTER NOTEBOOKS ON GRAHAM

JOSE NANDEZ
SHARCNET
UNIVERSITY OF WESTERN ONTARIO
JUPYTER NOTEBOOK

• Open-source web application

• allows you to create and share documents that contain live code, equations, visualizations and narrative text.

• Uses: data cleaning and transformation, data visualization, machine learning, etc.
• *pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for Python.
MATPLOTLIB

- Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
- Matplotlib can be used in Python scripts, jupyter notebook, web application servers, and four graphical user interface toolkits.
SKLEARN

• Simple and efficient tools for data mining and data analysis
• Built on NumPy, SciPy, and matplotlib
• Open source, commercially usable - BSD license
• Mathematical models:
  • Classification
  • Regression (example today)
  • Clustering
  • Dimensionality reduction
  • Model selection (example today)
  • Preprocessing
REFERENCES

- [http://jupyter.org/documentation.html](http://jupyter.org/documentation.html)
- [http://pandas.pydata.org/](http://pandas.pydata.org/)
- [https://matplotlib.org/](https://matplotlib.org/)
- Get the notebooks from: [https://github.com/JNandez/jupyter_sharcnet_graham](https://github.com/JNandez/jupyter_sharcnet_graham)