mistywest

Open Source for Data Scientists

Presented by Andreas Putz

28.11.2019

Github Repository

README.md

Quick introduction to Scikit-Learn

A small demo for the Vancouver Open Source Software event hosted by the Microsoft Garage.

Anaconda Setup

Download and install Anaconda

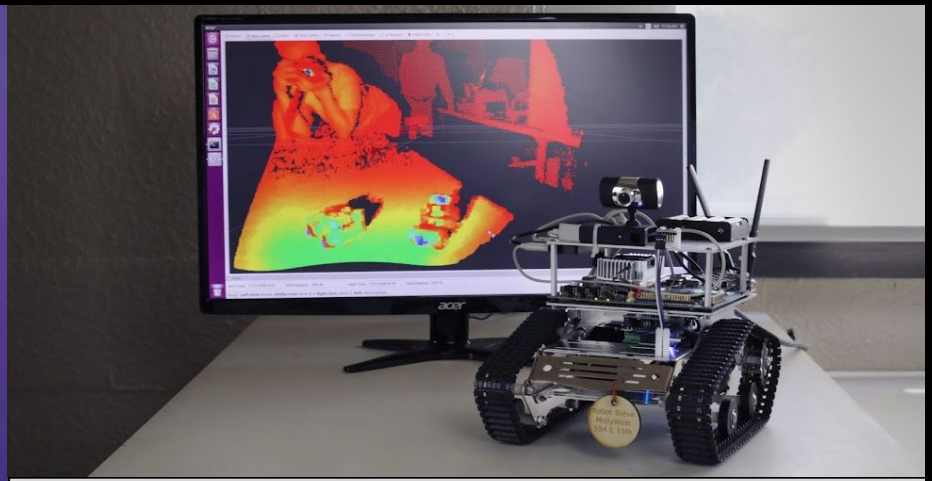
```
$ conda create --name MSGarage python=3.7
$ conda install jupyter scipy pandas matplotlib scikit-learn seaborn pydotplus scikit-image opencv tensorf
$ conda activate MSGarage
$ jupyter-notebook
```

https://github.com/putza/msgarage-sklearn

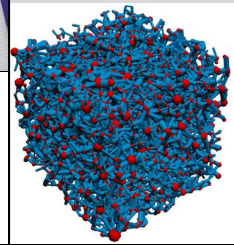
$ whoami

openpnm

An Open Source Pore Network Modeling Package

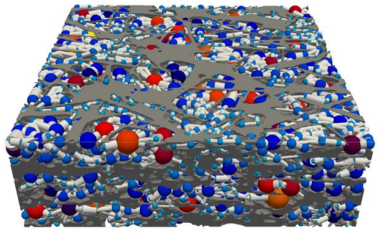


Import networks extracted using Maximal Ball, iMorph, 3DMA Rock, SNOW, and more.



Left: Berea sandstone network extracted using the Maximal Ball algorithm.

Right: Toray fibrous electrode extracted using the SNOW algorithm.



https://github.com/putza/msgarage-sklearn

Learn Data Science

Vancouver, BC

4,467 members · Public group

Organized by **Charles Iliya K.** and **17 others**

Share:

What we're about

Welcome to the Data Science LEARNING Group!

A meetup for people who want to LEARN Data Science as a group. Taking online courses together. Reading books together. Etc. Also with some hands-on workshops taught by 'experts'. It's also a place where you can ask others questions and for help.

Organisers

Charles Iliya K. and **17 others**

Message

Members (4,467)

See all

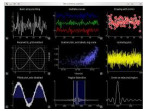
Open Source for Data Scientists

The Python Data Science Stack

Machine Learning

Image Processing

Data Visualization



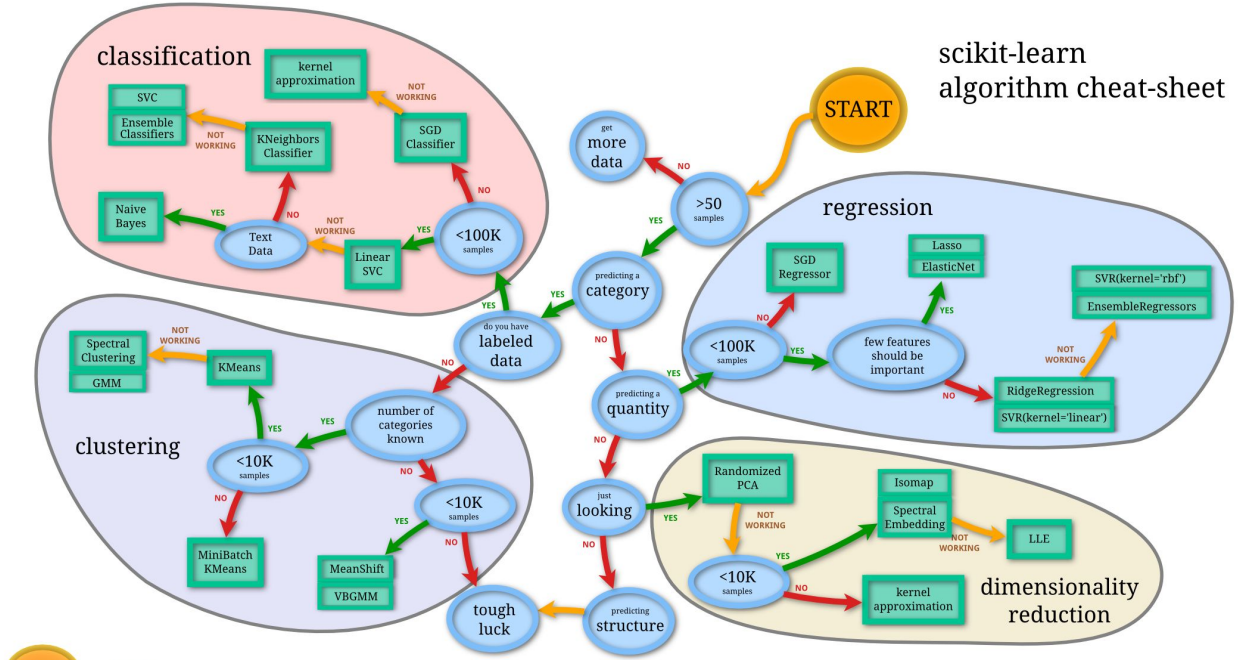
Data Exploration

Interpreters

Scikit-Learn

- Sits on top of Scipy, Pandas and Matplotlib
- Large collection of ML models
- Handles a lot of the logistics (e.g random splits of training, validation and testing data)
- Plays well with other packages

Scikit-Learn



Source: Scikit Learn https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Open Source for Data Scientists

Python In Virtual Environment

ANACONDA.

- Easy cross platform
- Rich ecosystem
- Easy to start
- Virtual environments (user/system)
- Cheat Sheet

- Download Anaconda
- Create a new environment
- Activate MSGarage
- Install packages
- Install your favourite editor (vi, ...)

```
$ conda create --name MSGarage python=3.7
$ conda install jupyter scipy pandas
matplotlib scikit-learn seaborn pydotplus
graphviz scikit-image opencv tensorflow
$ conda activate MSGarage

$ git checkout

$ jupyter-notebook
```

Python Virtual Environments

- Self contained
- Great for server environments

https://realpython.com/python-virtual-environments-a-primer/

Data Science Project With Cookie Cutter

```
├── LICENSE
├── Makefile           <- Makefile with commands like `make data` or `make train`
├── README.md          <- The top-level README for developers using this project.
├── data
│   ├── external       <- Data from third party sources.
│   ├── interim        <- Intermediate data that has been transformed.
│   ├── processed      <- The final, canonical data sets for modeling.
│   └── raw            <- The original, immutable data dump.
│
├── docs               <- A default Sphinx project; see sphinx-doc.org for details
│
├── models             <- Trained and serialized models, model predictions, or model summaries
│
├── notebooks          <- Jupyter notebooks. Naming convention is a number (for ordering),
│                         the creator's initials, and a short `-` delimited description, e.g.
│                         `1.0-jqp-initial-data-exploration`.
│
├── references         <- Data dictionaries, manuals, and all other explanatory materials.
│
├── reports            <- Generated analysis as HTML, PDF, LaTeX, etc.
│   └── figures        <- Generated graphics and figures to be used in reporting
│
├── requirements.txt   <- The requirements file for reproducing the analysis environment, e.g.
│                         generated with `pip freeze > requirements.txt`
│
├── setup.py           <- makes project pip installable (pip install -e .) so src can be imported
├── msglearn           <- Source code for use in this project.
│   ├── __init__.py    <- Makes src a Python module
│   │
│   ├── data           <- Scripts to download or generate data
│   │   └── make_dataset.py
│   │
│   ├── features       <- Scripts to turn raw data into features for modeling
│   │   └── build_features.py
│   │
│   ├── models         <- Scripts to train models and then use trained models to make
│   │   │                 predictions
│   │   ├── predict_model.py
│   │   └── train_model.py
│   │
```

https://drivendata.github.io/cookiecutter-data-science/

Open Source for Data Scientists

DEMO

Thank You!

mistywest

contact@mistywest.com

mistywest.com

mistywestyvr

mistywestconsulting

mistywest