Getting started with Anaconda and Jupyter Notebooks
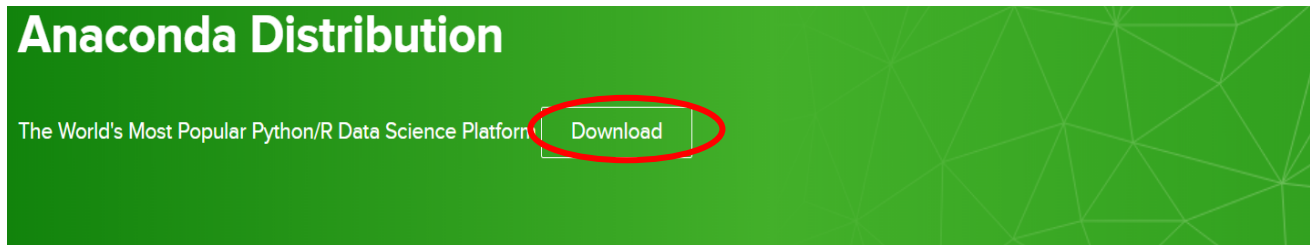
1. Go to Anaconda website: https://www.anaconda.com/distribution/
   Download Anaconda Python

Anaconda Distribution

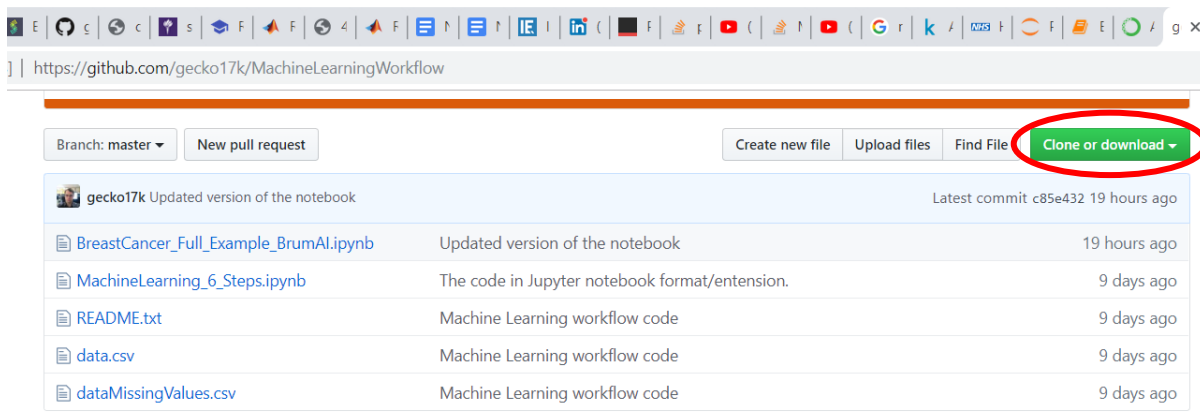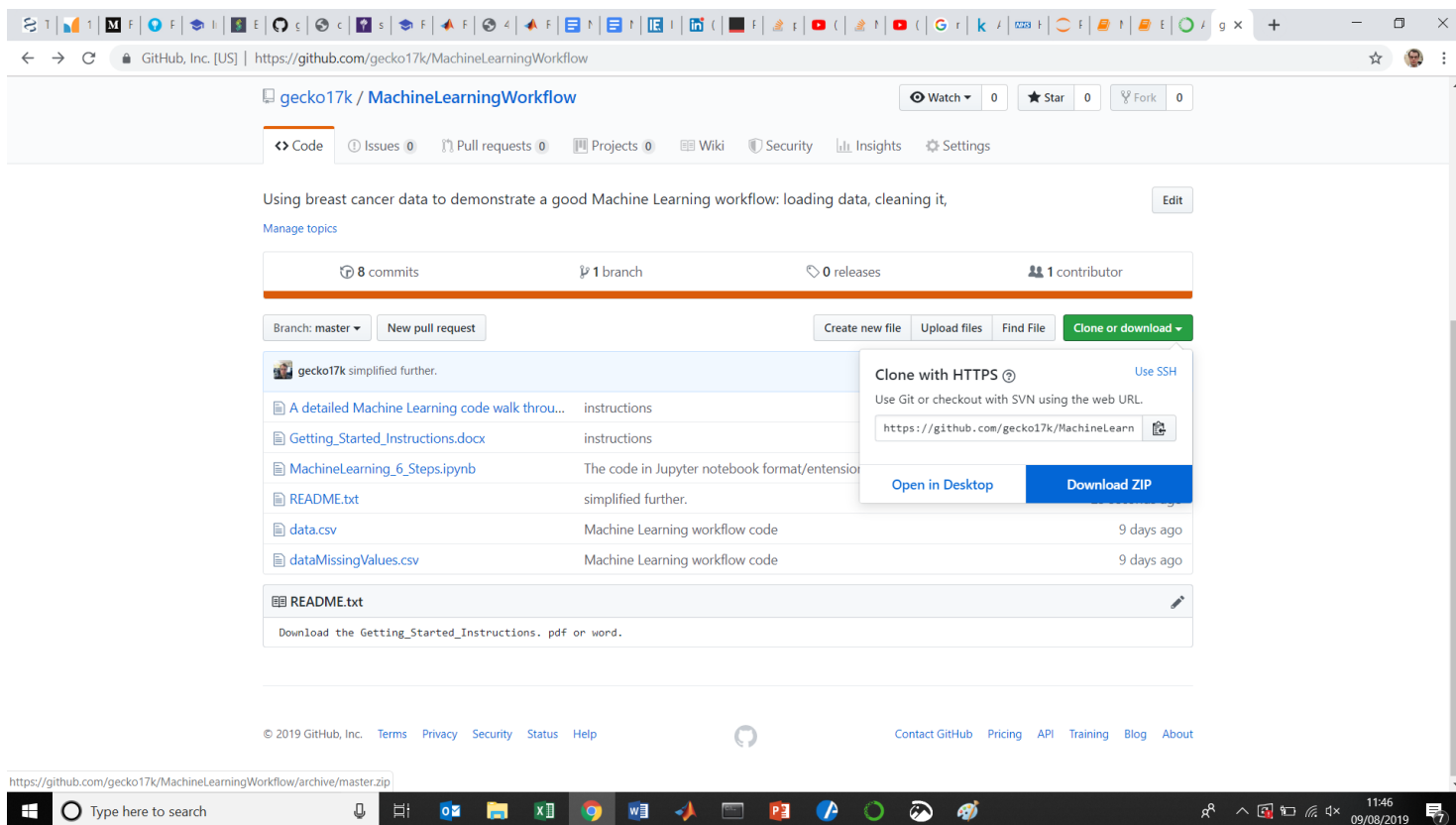The World's Most Popular Python/R Data Science Platform — **Download**

The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 15 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with Conda
- Develop and train machine learning and deep learning models with scikit-learn, TensorFlow, and Theano
- Analyze data with scalability and performance with Dask, NumPy, pandas, and Numba
- Visualize results with Matplotlib, Bokeh, Datashader, and Holoviews

Windows | macOS | Linux
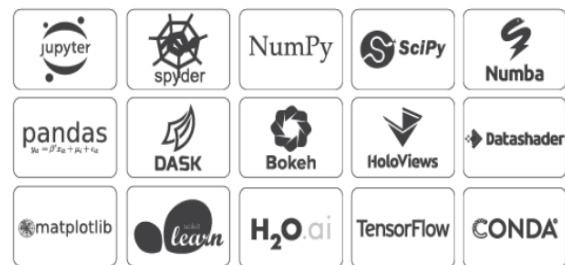
1. Go to GitHub, get the repo: https://github.com/gecko17k/MachineLearningWorkflow
   Clone or download the files shown below, easiest to get the ZIP file.

https://github.com/gecko17k/MachineLearningWorkflow

Branch: master ▾	New pull request		Create new file	Upload files	Find File	Clone or download ▾

gecko17k Updated version of the notebook — Latest commit c85e432 19 hours ago

| File	Description	Time
BreastCancer_Full_Example_BrumAI.ipynb	Updated version of the notebook	19 hours ago
MachineLearning_6_Steps.ipynb	The code in Jupyter notebook format/entension.	9 days ago
README.txt	Machine Learning workflow code	9 days ago
data.csv	Machine Learning workflow code	9 days ago
dataMissingValues.csv	Machine Learning workflow code	9 days ago

2. Unzip/Extract

Once you unzip, the relevant folder will be called "MachineLearningWorkflow-master".
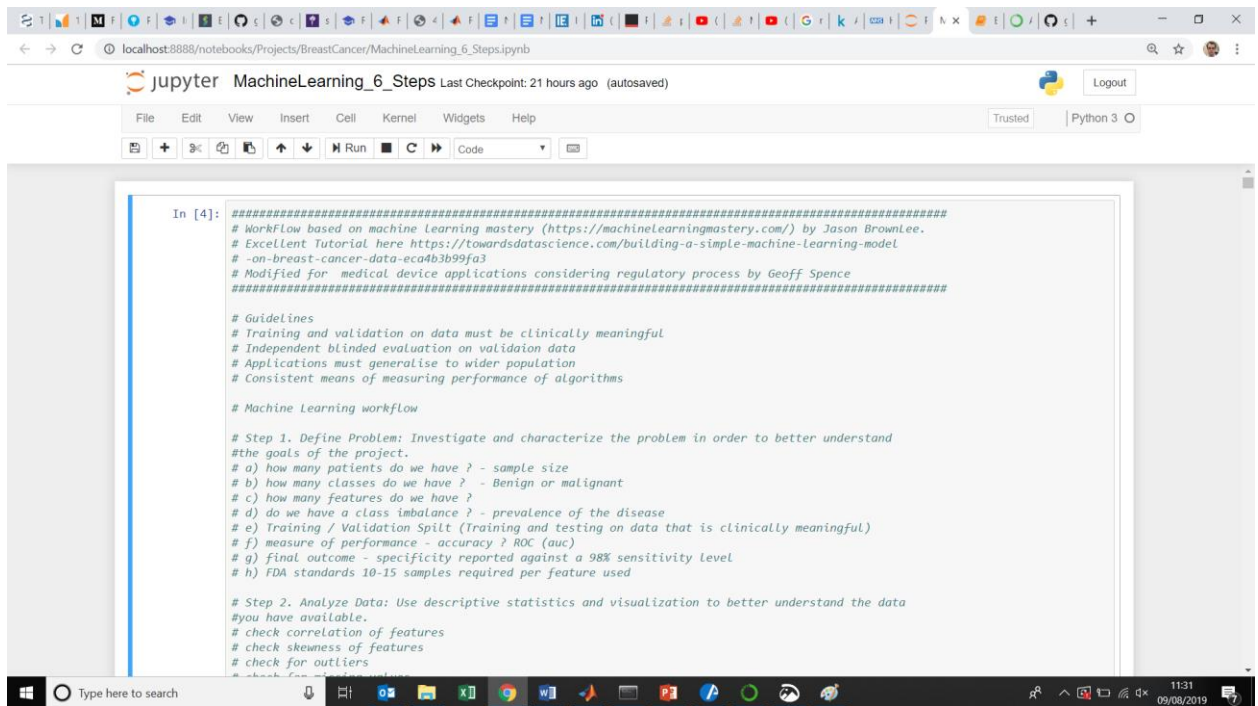
3. Place in a single folder on your laptop, e.g. C:\Users\Name\Projects\BreastCancer or a folder called "BreastCancer".

4. Open Anaconda Navigator
   Launch Jupyter Notebook
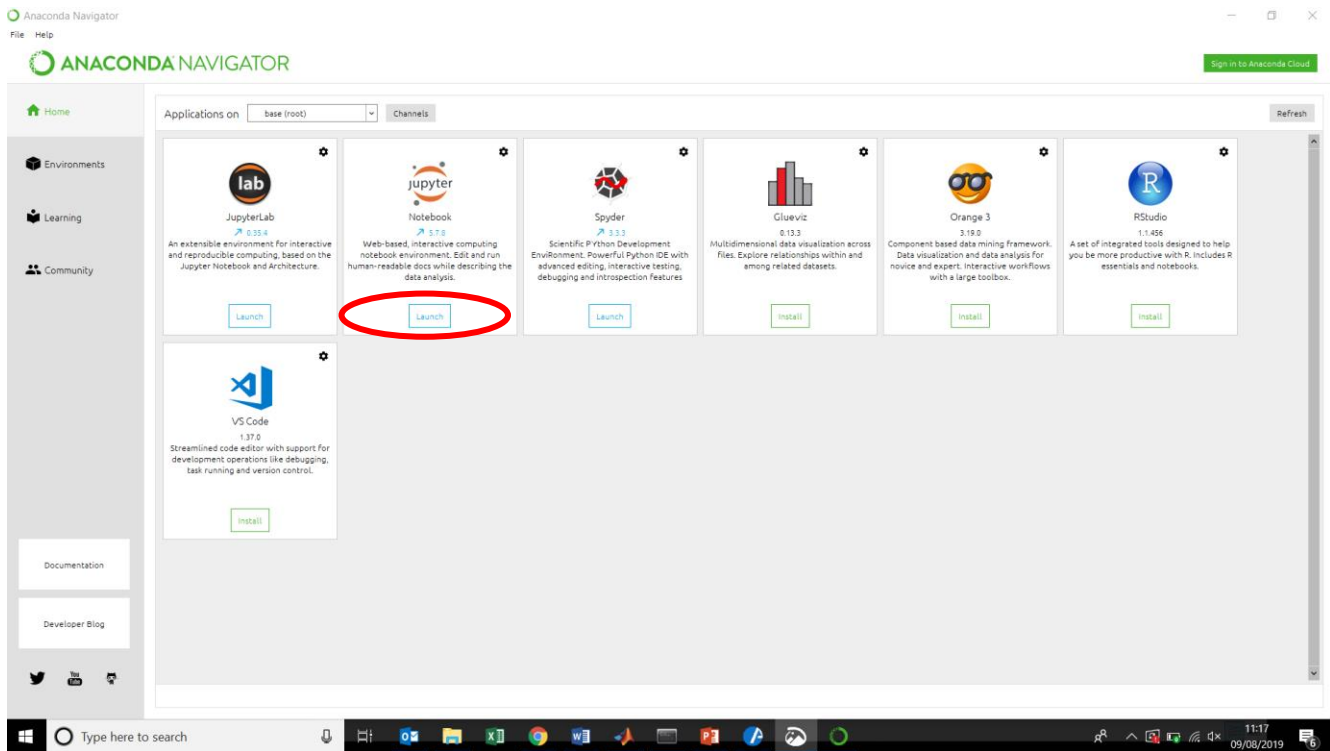
5. Jupyter opens in your browser.
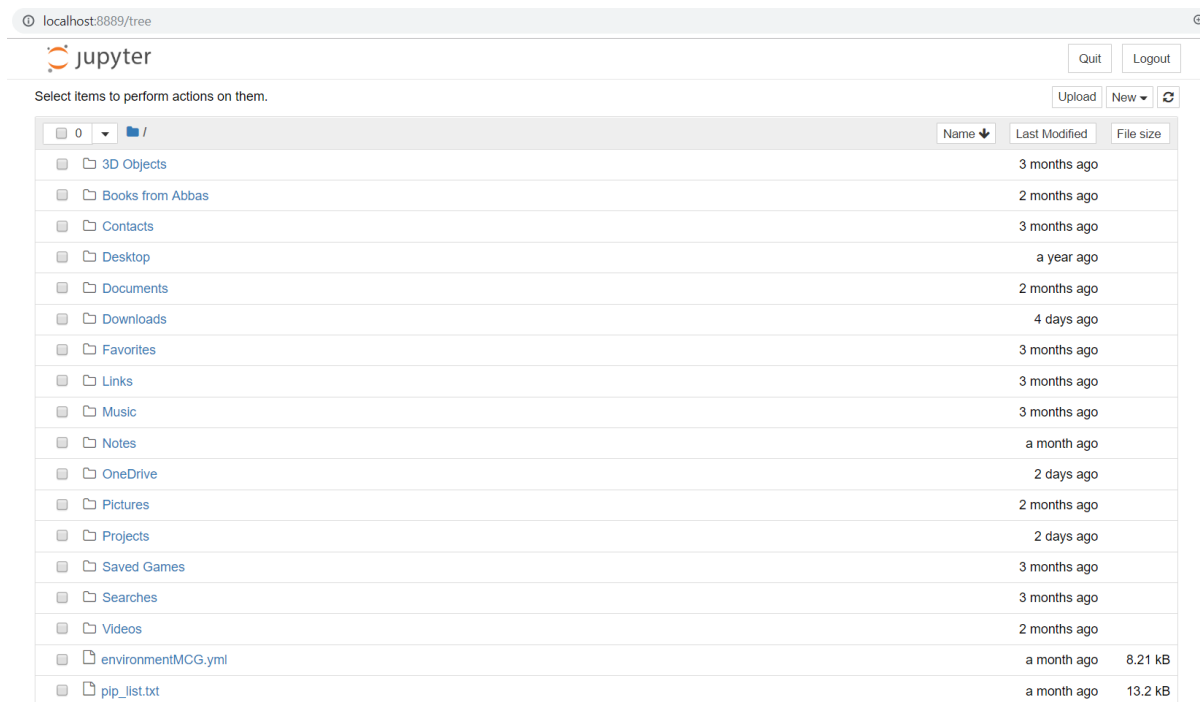
6. Navigate to where you put the files from GitHub and open up "-MachineLearning_6_Steps.ipynb".

Jupyter MachineLearning_6_Steps Last Checkpoint: 21 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 O

```
In [4]: #######################################################################################
        # WorkFlow based on machine learning mastery (https://machinelearningmastery.com/) by Jason BrownLee.
        # Excellent Tutorial here https://towardsdatascience.com/building-a-simple-machine-learning-model
        # -on-breast-cancer-data-eca4b3b99fa3
        # Modified for  medical device applications considering regulatory process by Geoff Spence
        #######################################################################################

        # Guidelines
        # Training and validation on data must be clinically meaningful
        # Independent blinded evaluation on validaion data
        # Applications must generalise to wider population
        # Consistent means of measuring performance of algorithms

        # Machine Learning workflow

        # Step 1. Define Problem: Investigate and characterize the problem in order to better understand
        #the goals of the project.
        # a) how many patients do we have ? - sample size
        # b) how many classes do we have ?  - Benign or malignant
        # c) how many features do we have ?
        # d) do we have a class imbalance ? - prevalence of the disease
        # e) Training / Validation Spilt (Training and testing on data that is clinically meaningful)
        # f) measure of performance - accuracy ? ROC (auc)
        # g) final outcome - specificity reported against a 98% sensitivity level
        # h) FDA standards 10-15 samples required per feature used

        # Step 2. Analyze Data: Use descriptive statistics and visualization to better understand the data
        #you have available.
        # check correlation of features
        # check skewness of features
        # check for outliers
```

Type here to search

11:31
09/08/2019

We are now ready to start.