

Welcome to the Python Party!

Winter 2019

Introduction

Welcome to an unplanned, poorly structured, romp through pythonic machine learning. **This is not a course.** It is more like a guided tour where the tour guide is constantly checking his phone for directions.

The objective of these tutorials is to facilitate independent learning amongst yourselves through guided material and assignments. Each week we'll have an opportunity to come back together and review via practice. Each tutorial will be focused doing some coding. We will be relying heavily on some materials pulled from Coursera, as well as some freely available resources. So consider this a framework and a resource, more than a fully structured curriculum.

One final note: this set of documents is being pulled together on the fly in my free time. The coding assignments will in most cases be modified from a different language to be accessible in Python3. There will be typos. There will be bugs. My apologies in advance.

1 Planned Organization

1.1 Python3

Although there may be moments where other languages (specifically MATLAB) are mentioned in the available texts or video lectures we will be using exclusively Python3. The differences between Python2 and Python3 are subtle, but it will serve you best to be consistent. Excellent references can be found on the language and the packages we will use with it.

1.2 Jupyter Notebooks

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Not only can we do everything we could in a stand-alone python code, but it is interactive and easy to process code step by step. This way, you can find and correct errors much easier. It also allows for us to put plain text alongside our code, so that we can communicate our purpose better. Some would argue your scientific life and collaborations would be more productive if you could email around jupyter notebooks instead of clunky origin files with word-salad emails.

1.3 Github

Github is a means of sharing and collaborating on large coding projects. The Computational Caucus inside the group makes regular use of a github for sharing files, and include public links on the publications that make use of any in-house programs.

The important thing here, is I will be keeping all of the source and some `.pdfs` for the tutorials in a repository on that github. It'll be your first stop each week to get started on new material. There may also use the github to raise issues with the assignments, and get help from your peers.

1.4 Dropbox

Github isn't ideal for hosting large files (like video lectures or powerpoints), so I will include links to Dropbox folders in the github files. The files will only be hosted on Dropbox for the week they are relevant, so get 'em while they're hot. This will likely be your second stop each week to get going on new material.

1.5 What's in a week

The tentative plan, is to have a set of video lectures assigned during the week. The lectures are very brief and can be accomplished in short spurts or at 1.5x speed. These may have accompanying documents as optional reading/review.

Early on they will almost exclusively be from Andrew Ng's courses. I will suggest registering for his Machine Learning Coursera course. The course is free to partake in, and has some quizzes and readings that can reinforce what you learn in the videos. Its assignments will be in MATLAB though, which will not be helpful as we dive into Deep Learning. A link for the course is here: <https://www.coursera.org/learn/machine-learning/home/info>

Following a touch of independent study on your own time over the week, we will meet up as a group. These tutorials will be almost exclusively practical, with a guided programming assignment using Jupyter Notebooks. The goal is to structure them in chunks, so that a topic can be reintroduced, and you can work with your neighbors to complete the necessary code. Naturally everyone will work at their own pace; however, the hope here is to encounter struggles together. I will also be there — with potentially others — to help nudge groups along who are stuck.

Then just like that you will have learned a new thing, practiced your python, enjoyed some problem solving, and be handed an updated Github repository with a new week's plan. How exciting!

1.6 Email

Last bit of organization: make sure I have your email on my spreadsheet. As this whole endeavor is happening on the fly, I'm sure there will be changes as we go, so it'd be ideal if I can contact the group.

We could also make a Slack channel if people are interested.

2 Get going with software

2.1 Anaconda

We will be using everyone's favorite data-science platform: Anaconda. This is a local installation of Python with some bells, whistles, and a graphical user interface. It will allow you to run Python without administration rights on your computer, and make installing more bells and whistles much easier.

1. Head to <https://www.anaconda.com/download/>
2. It should automatically bring up the download page for your operating system. Click the big green "Download" button beneath "Python 3.7 Version *"
3. Open the downloaded installer and follow the instructions to get anaconda installed.
4. Once installed open up the Anaconda navigator and click on the **Environments** tab.
5. Click **+Create** and name an environment ("ML" would suffice).
6. You will see all of the packages available on the right. Change the dropdown menu currently on Installed, to Not Installed.
7. Using the search bar, search and check off the following packages (typed exactly):
 - jupyter
 - numpy
 - tensorflow
 - keras
 - scikit-learn
 - pandas
 - imageio
 - opencv
 - matplotlib
8. Click **Apply**

This should give you all of the necessary packages for the work we will do.

2.2 Jupyter

Navigate back to the **Home** tab. Underneath the orange and grey jupyter icon, click **Launch**. This will open up a web browser, where you can navigate files and run jupyter notebooks. Lets leave this open, but ignore it for a few moments...

3 Week 0 Tutorial

Now that we have the necessary software, time to do some coding! If you haven't already, head to <https://github.com/CooperComputationalCaucus/ml-tutes>. This is where I will be keeping the necessary course material for each week. Open if you are familiar with git, pull the repository, otherwise download the zip. Unzip the file to your preferred location, and then head back to the browser running jupyter, and navigate to the folder you just unzipped. Open the file labeled `python3_crash_course.ipynb`.

4 Week 1 Material

Over the next week please read and watch the following:

- Lecture 1
 - What is machine learning
 - Supervised learning
 - Unsupervised learning
- Lecture 2
 - Model representation
 - Cost function
 - Cost function intuition I+II
 - Gradient descent
 - Gradient descent intuition
 - Gradient descent linear regression
- *Optional* Linear algebra review
 - Lecture 3
 - Matrices and vectors
 - Addition and Scalar Multiplication
 - Matrix vector multiplication
 - Matrix matrix multiplication
 - Matrix multiplication properties
 - Inverse and transpose
- Lecture 4
 - Multiple features
 - Gradient descent for multiple variables
 - Gradient descent in practice I+II
 - Features and polynomial regression

- Normal equation
- Normal equation noninvertability

This corresponds to Week 1 and Week 2 of the Coursera course. Files for these can be found here:

<https://www.dropbox.com/sh/8kyk7hxvy8cckkv/AAAqspCXNYTMeGZonyzv2oHua?dl=0>.