

A Code-First Introduction to NLP

A New TWiML x fast.ai Study Group
Based on the course
created by Rachel Thomas,
co-founder of fast.ai

Joseph Catanzarite
14 Dec 2019

Course Overview

- Excited to be here, hope you are too!
- How to make Learning Online work
- Course prerequisites
- Compute infrastructure, running course notebooks
- “Top-down approach” and how to make it work for you
- Topics to be covered
- Optional assignment: write a blog post

Learning Online

- Make our meetup a classroom environment
 - Ask questions, offer comments
 - Often questions and comments help others learn!
 - Please make sure you are muted! Unless you are talking.
 - Stream video, if you have the bandwidth, so we can see each other!
- Take advantage of the fastai community
 - Post your questions and comments on the fastai Forum
 - If can usefully answer someone's question, consider posting it!

Course Prerequisites

- Familiarity with
 - working with data in Python
 - machine learning concepts, such as training and test sets
- Some experience with PyTorch and neural networks is helpful but not necessary

Set up your compute infrastructure

1. Download the Anaconda Python 3.7 distribution at

`https://www.anaconda.com/distribution/#download-section`

2. Install Anaconda by following the instructions at

`https://docs.anaconda.com/anaconda/install/`

2. Open an anaconda shell terminal (choose the Anaconda Prompt from the Start Menu in Windows)

3. Create an environment for fastai (See <https://uoa-eresearch.github.io/eresearch-cookbook/recipe/2014/11/20/conda/>),

then activate that environment:

```
conda create --name fastai
```

```
activate environment
```

4. Create a fastai directory and Install the fastai library:

```
mkdir fastai
```

```
conda install -c pytorch -c fastai fastai
```

5. Make a local copy of the fastai course-nlp repository:

```
cd fastai
```

```
git clone https://github.com/fastai/course-nlp.git
```

Running a course notebook

1. Open an anaconda shell, with the Anaconda Prompt

2. Activate the fastai environment

```
activate fastai
```

2. Go to your local copy of the course repository

```
cd fastai/course-nlp/
```

3. Run the command

```
jupyter notebook
```

A new browser tab will open, with links to the course notebooks in the repository

4. Clicking on a selected notebook link opens that notebook in a separate browser tab

5. We'll cover the notebook `2-svd-nmf-topic-modeling` next week. Click on it and verify that you can run the code blocks within it.

Top down approach

- Philosophy is to start “doing” as soon as possible
- Potential problem: information overload!
 - Notebooks, videos, references, blogs, papers, code packages
 - Can be overwhelming!
- Solution: keep to a straight path, limit your scope
 - Focus on the videos and the course notebooks, avoid distraction
 - Don't feel you have to follow up all the peripheral information!
- If you're already familiar with much of the material, and you have time, then by all means do check out the suggested references.

Topics covered

- Topic Modeling with NMF (non-negative matrix factorization) and SVD (Singular Value Decomposition)
- Sentiment classification with Naive Bayes, Logistic regression, and ngrams
- Regex and tokenization
- Language modeling & sentiment classification with deep learning
- Translation with RNNs
- Translation with the Transformer architecture
- Bias & ethics in NLP
- See [notebook0](#) for a complete list, and references

Optional assignment

- Consider this quote, from Albert Einstein: “You don’t **really** understand anything unless you can explain it to your grandmother”
- Write a blog post about something you’ve learned in the class
- Post it to the Forum, on Medium, or another online venue

First video

- Course covers
 - Topic Modeling
 - Sentiment Classification
 - Language Modeling
 - Translation
- Generally I rely on you to watch videos, then I pick some parts of the video to cover in depth
- [My github repo](#)
- [NLP Study Group repo](#) contains original and annotated course notebooks and other course materials (such as this document!)
- [Peter Norvig spell-checker notebook](#)