Tutorial 9 (15 pt.) Latent Dirichlet Allocation for topic clustering using python

Submit Assignment

Due No Due Date **Points** 15 **Submitting** a file upload
File Types pdf, txt, zip, gzip, ipynb, and py
Available Jan 14 at 12am - Mar 13 at 11:59pm about 2 months

This tutorial is advanced. It is recommended that one complete Olga's Varational Inference tutorial 11 before this.

One should read the paper that is included but the main points are summarized in the tutorial.

The tutorial should take 6-10 hours if you have the right knowledge coming in.

The grading can earn you 1,9,11,13 or 15 points depending on how many things you show us you learned and understood. To pass at all you must show knowledge around the underlying problem, the model (Figure 1 in the paper), and the variational approach used to solve it.

So one seems to need to run this in python 3 and in jupyter notebook.

If you do not have jupyter notebook installed see:

http://jupyter.readthedocs.io/en/latest/install.html\
(http://jupyter.readthedocs.io/en/latest/install.html%5C)

You need to get it set up on mu laptop I did:

apt-get install python3-numpy
apt-get install python3-scipy
apt-get install python3-sklearn

On the ubuntu computers at KTH you can for example do:

```
python3 -m pip install --user numpy
```

Tutorial is the LDATutorial.ipynb file in:

PGM_LDA_Tutorial.tar.gz

Go to the PGM_LDA_Tutorial folder after unziping it.

Type jupyter notebook to launch the Jupyter Notebook App The notebook interface will appear in a new browser window or tab. click on the LDATutorial.ipynb file to open it

Note that there is a kernel menu in Jupyter that you can select the python 3

When you get to code you can edit it directly and then run it by clicking the run button while the code box is active. Output appears then below the code box.

Note that you must implement figure 6 from the paper (VI Parameter Estimation) for the E-step and the 'big psi' there is diGamma() in the code

Also note that 'converged' means the sum of the absolute values of the change in gamma[d][i] summed over i is less than eta

That is stated lower down in the instructions but I just sort of jumped in early and was confused.