

## SI 630: Homework 4

Yashaswini Joshi (Unique Name: yjoshi)

### 2. Implementing a Gibbs Sampler (80 points)

- 2.1 Changing topic assignments / counts
- 2.2 Computing the sampling distribution

**Ans:** Implementation can be found in the `lda.py` file.

### 3 Running on Your Code Real Data (10 Points)

**Ans:** `my-topics.txt` file is uploaded in canvas.

The time taken for the model to finish 1000 iterations in seconds is 286.

### 4 Comparing your model (10 points)

1. Whether each model found reasonable topics – do the top words for a topic make it seem like a coherent theme?

**Ans:** There exists coherent theme. Topic 0 of GenSim has nuclear, military, medicine and all which tells it is related to something regarding military operation. But it also consists of monk and born which seems unrelatable to the topic. But in case of Mallet this is not seen much. All the words seem relatable and coherent.

2. Whether each model found the same kind of topics

**Ans:** Mallet model's output had similar kind of topics compared to that of GenSim's.

3. How the models differed in speed (be sure to report the times) and whether this seems related to topic quality

**Ans:** Mallet took more time compared to GenSim. GenSim took 5.02 seconds to run on wiki data while Mallet took 6.8 seconds. But the output of Mallet is more relatable compared to that of GenSim's output.