Advanced Probabilistic Machine Learning and Applications

Pablo Sánchez and Isabel Valera

1 Tutorial 3: Bayesian Mixture Model (BMM)+ Gibbs sampling

In this tutorial we will continue working with the CMM and Twitter data-set presented in Tutorial 2. We will use different versions of the Gibbs sampling algorithm to find the posterior distribution of the cluster assignments $\{z_n\}_{n=1}^N$ and model parameters $(\pi, \{\boldsymbol{\theta}_k\}_{k=1}^K)$.

Introduction

Notation: Through this document we will use the following notation:

- K: number of mixture components, i.e., we interpret them as topics/clusters.
- N: number of documents, i.e., tweets.
- I: dictionary
- $|I|$: number of words in I.
- $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^K$: set of likelihood parameters.
- $\mathbf{x}_n \in \mathbb{R}^{W_n}$: n-th document with length (i.e., number of words) W_n.
- $\boldsymbol{X} = \{\mathbf{x}_n\}_{n=1}^N$: set of all documents.
- $\boldsymbol{X}_{-n} = \{\mathbf{x}_i | i \neq n\}_{i=1}^N$: set of all documents except for $\mathbf{x}_n$.
- z_n:component assignment variable of document $\mathbf{x}_n$.
- $\boldsymbol{Z} = \{z_n\}_{n=1}^N$: set of all component assignment variables.
- $\boldsymbol{Z}_{-n} = \{\mathbf{z}_i | i \neq n\}_{i=1}^N$: set of all component assignment variables except for z_n.

Summary of Generative Model: We will work with the following Bayesian Mixture Model

$$p(\boldsymbol{X}, \boldsymbol{Z}, \pi, \boldsymbol{\Theta}) = p(\pi|\boldsymbol{\alpha})p(\boldsymbol{\Theta}|\boldsymbol{\gamma})\prod_{n=1}^{N}[p(z_n|\boldsymbol{\pi})p(\mathbf{x}_n|z_n, \boldsymbol{\Theta})]$$

The conjugate prior for the categorical distribution is the Dirichlet distribution. Therefore, we define the prior distribution for $\boldsymbol{\pi}$ and $\boldsymbol{\theta}_k$ for all k as Dirichlet distributions with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ respectively. Notice the prior distributions for each $\boldsymbol{\theta}_k$ share the same set of parameters.

$$p(\pi|\boldsymbol{\alpha}) = \mathrm{Dir}(\pi|\boldsymbol{\alpha}) \quad p(\boldsymbol{\Theta}|\boldsymbol{\gamma}) = \prod_{k=1}^{K}\mathrm{Dir}(\boldsymbol{\theta}_k|\boldsymbol{\gamma}) \quad p(z_n|\boldsymbol{\pi}) = \mathrm{Cat}(z_n|\boldsymbol{\pi}) \quad p(\mathbf{x}_n|z_n, \boldsymbol{\Theta}) = \prod_{j=1}^{W_n}\mathrm{Cat}(x_{nj}|\boldsymbol{\theta}_{z_n})$$

Submission: Copy the Jupyter notebook available in the Github repository https://github.com/APMLA/apmla_material/tree/master/L3 and complete the exercises proposed below. You will need to submit electronically the complete version of the Jupyter (together with the future exercises for Block I) by December 13th.

Exercise 1: Derive the Gibbs sampling Algorithm for the CMM

Given the dataset and the probabilistic model described in the previous section, complete the following tasks by hand or in latex:

1. **Algorithm 1:** Use the Gibbs sampling algorithm to approximate (using samples) the posterior distribution $p(\pi, Z, \Theta | X)$. Derive the posterior.

Algorithm 1: Gibbs sampling algorithm

Initialize cluster assignments Z and the model parameters π, Θ;
while *not converged* **do**
  Sample $\pi \sim p(\pi | X, Z, \Theta) = p(\pi | Z)$;
  for $k = 1, \ldots, K$ **do**
    Sample $\theta_k \sim p(\theta_k | X, Z, \pi) = p(\theta_k | X, Z)$;
  end
  for $n = 1, \ldots, N$ **do**
    Sample $z_n \sim p(z_n | X, Z_{-n}, \pi, \Theta) = p(z_n | \mathbf{x}_n, \pi, \Theta)$;
  end
end

2. **Algorithm 2:** Use the (collapsed) Gibbs sampling algorithm to approximate (using samples) the posterior distribution $p(Z, \Theta | X)$. Derive the posterior.

Algorithm 2: π collapsed Gibbs sampling algorithm

Initialize cluster assignments Z and the model parameters Θ;
while *not converged* **do**
  for $k = 1, \ldots, K$ **do**
    Sample $\theta_k \sim p(\theta_k | X, Z)$;
  end
  for $n = 1, \ldots, N$ **do**
    Sample $z_n \sim p(z_n | X, Z_{-n}, \Theta) = p(z_n | \mathbf{x}_n, Z_{-n}, \Theta)$;
  end
end

3. **Algorithm 3:** Use the (collapsed) Gibbs sampling algorithm to approximate (using samples) the posterior distribution $p(Z | X)$. Derive the posterior.

Algorithm 3: π, Θ collapsed Gibbs sampling algorithm

Initialize cluster assignments Z;
while *not converged* **do**
  for $n = 1, \ldots, N$ **do**
    Sample $z_n \sim p(z_n | X, Z_{-n})$;
  end
end

Exercise 2: Algorithms implementation in Python

1. Implement in Python the three Gibbs samplers derived in Exercise 1.
2. For each of the three Gibbs samplers and $K = 4$, get samples from the hidden variables after different burn-in periods $\tau_{\text{burn-in}} \in \{20, 200, 500\}$ [1]. Do you think the Gibbs sampler has converged (i.e the samples are from the target posterior distribution)?
3. Let consider the log-likelihood as the measure of convergence and $K = 5$. For each of the three Gibbs samplers, show the evolution of the log-likelihood per iteration until convergence.

[1] See L3 notes for more information about the burn-in period.

Then, take a sample from all the hidden variables after convergence and show i) the 10 most representative words for each topic using a cloud of words and ii) the 10 most relevant documents for each topic.

Hint: $\log p(\boldsymbol{X}|\boldsymbol{\Theta}, \pi, \boldsymbol{Z})$