Background on Labeled Latent Dirichlet Allocation

1 Latent Dirichlet allocation

Latent Dirichlet Allocation (LDA) is a parametric probabilistic generative model, sometimes referred to as "topic model". It is used for discovering the main themes that pervade a large, unstructured collection of documents. The discovered topics reflect themes associated with the documents. We assume that you have some knowledge of what LDA is. Consult [1] for a brief overview and [2] if you are interested to know more details. You are not required to know the details explained in [2].

2 Labeled LDA

2.1 Intuition

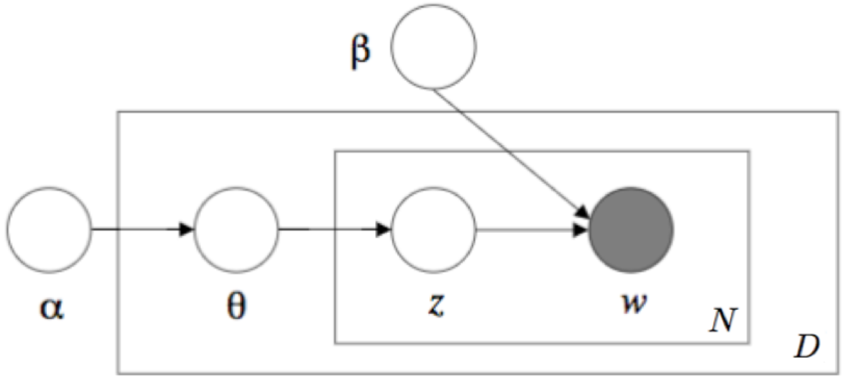
Briefly, Labeled LDA (LLDA) is a parametric probabilistic model used to learn topics for documents that are associated with several labels each. It is an extension of LDA where a one-to-one relationship is asserted between the learned topics and labels. Labeled LDA models each document as a mixture of underlying themes, assuming a generative process where each word is generated from a topic. Unlike LDA, LLDA incorporates supervision by constraining the topic model to use only those topics that correspond to a document's (observed) label set.

2.2 Generative model

A generative model describes the process which we hypothesize generated the observed data. We use the following notation:

- K: the number of topics in the entire collection of patient charts. In both LDA and LLDA, this is a user specified parameter. Because the user has to specify a finite number of parameters assumed to exist in the model, we say that LLDA and LDA are parametric models
- V: is the total number of words in the entire collection of patient charts, while N is the number of words within a patient chart.
- D: is the number of observations in the dataset; that is the number of documents or in our case the number of ICU stays.
- β: is a multinomial distribution for each $k \in K$. It can be thought of as a matrix of K columns and V rows. Every column specifies the word distributions for a given topic k. Cell k, v reflects how relevant word v is for topic k, extremely relevant words will have high $\beta_{k,v}$. It is common to arrange the words according to their relevance (measured by the magnitude of $\beta_{k,v}$ to understand the concepts that the topics reflect.
- θ: Can be thought of as a matrix of D rows and K columns. Each row tells us how relevant the K topics are for the corresponding patient.
- Λ: This is only available for LLDA not LDA. This is a binary matrix of D rows and K columns encoding whether or not a given patient had the corresponding label. Λ limits the topics that a patient can belong to. Unlike LDA, LLDA does not allow the patient specific topics to be drawn from the entire θ vector, but only from the topics which he or she has been labelled with.

Figure 1: LDA plate diagram



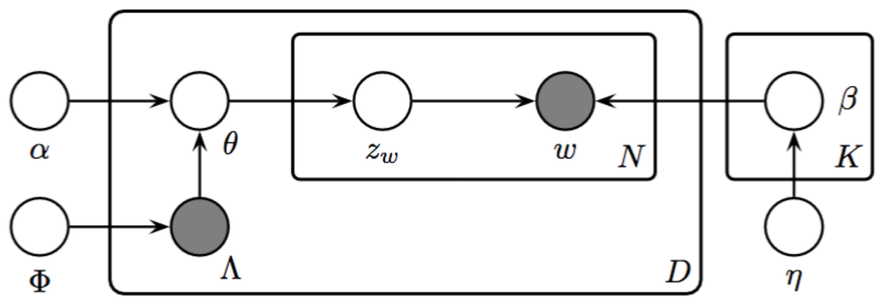
Similar to LDA, LLDA takes as an input a number of K topics specified by the user. The generative process proceeds by drawing a multinomial distribution β_k for every $k \in K$ using a Dirichlet prior η. Unlike LDA, LLDA does not draw the patient-chart specific topic mixtures from all K topics but rather limits the draws to the labels that the patient has. Mechanically, we do so by mapping the α parameter from a single vector (as is done in LDA) to a matrix with one entry for each patient. The entries corresponding to topics which a patient does not have are "turned off" when they are multiplied by a projection matrix L encoding the patient labels. Figure 1 shows the plate diagram corresponding to LDA while figure 2 shows the plate diagram for LLDA. Grey nodes denote observed variables while white ones denote latent ones. Specifically, the generative model for LDA proceeds as follows:

1. For each topic $k \in \{1, ..., K\}$: // These are global parameters not patient specific but topic specific.
   - Draw "word relevance to topic k": $\beta_k \sim \text{Dirichlet}(\eta)$
2. For each document $d \in \{1, ..., D\}$ // These are patient specific specific draws
   - For each topic $k \in \{1, ..., K\}$:
     - For each word w in the patient chart, i.e., $w \in \{1, ..., N_w\}$:
       - Draw a binary value z_w denoting whether or not w belongs to *any* topic
       - Draw a word w based on the relevance to the topic drawn in the previous step

This process is slightly different from that of LLDA's, which is specified as follows:

1. For each topic $k \in \{1, ..., K\}$: // These are global parameters not patient specific but topic specific.
   - Draw "word relevance to topic k": $\beta_k \sim \text{Dirichlet}(\eta)$
2. For each document $d \in \{1, ..., D\}$ // These are patient specific specific draws
   - For each topic $k \in \{1, ..., K\}$:
     - Draw binary patient labels $\Lambda^{(d)} \sim \text{Bernoulli}(\Phi_k)$
       - For each word w in the patient chart, i.e., $w \in \{1, ..., N_w\}$:
         - Draw a binary value z_w denoting whether or not w belongs to a topic, limiting the possible topics to draw from to the subset of labels the patient possesses
         - Draw a word w based on the relevance to the topic drawn in the previous step

Figure 2: LLDA plate diagram



2.3 Inference

After carefully outlining the generative process, we need to learn the different parameters that govern the statistical draws and the posterior distributions of the latent variables in the model. In order to do so, we typically resort to approximate inference to learn the parameters of LDA and LLDA. Mallet, the software you will use for this problem set, incorporates a sampling technique. Variational inference methods can also be used to learn the parameters. Inference methods are beyond the scope of this problem set (but tremendously interesting and important to understand!).

References

[1] David Blei, "Probabilistic topic models". *Communications of the ACM*. April 2012. `http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf`

[2] David Blei, Andrew Ng and Michael Jordan. "Latent Dirichlet Allocation". *JMLR*. 2003. `http://http://www.cs.columbia.edu/~blei/papers/BleiNgJordan2003.pdf`

[3] Daniel Ramage, David Hall, Ramesh Nallapati, Christopher D. Manning. "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora". *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. 2009. `https://nlp.stanford.edu/pubs/llda-emnlp09.pdf`