Latent Dirichlet Allocation

Jérôme DOCKÈS, Pascal LU

École Normale Supérieure de Cachan — December 29, 2015

Objectives

Lorem ipsum dolor sit amet, consectetur, nunc tellus pulvinar tortor, commodo eleifend risus arcu sed odio:

- Mollis dignissim, magna augue tincidunt dolor, interdum vestibulum urna
- Sed aliquet luctus lectus, eget aliquet leo ullamcorper consequat. Vivamus eros sem, iaculis ut euismod non, sollicitudin vel orci.
- Nascetur ridiculus mus.
- Euismod non erat. Nam ultricies pellentesque nunc, ultrices volutpat nisl ultrices a.

Presentation of the model

Lorem ipsum dolor **sit amet**, consectetur adipiscing elit. Sed commodo molestie porta. Sed ultrices scelerisque sapien ac commodo. Donec ut volutpat elit. Sed laoreet accumsan mattis. Integer sapien tellus, auctor ac blandit eget, sollicitudin vitae lorem. Praesent dictum tempor pulvinar. Suspendisse potenti. Sed tincidunt varius ipsum, et porta nulla suscipit et. Etiam congue bibendum felis, ac dictum augue cursus a. **Donec** magna eros, iaculis sit amet placerat quis, laoreet id est. In ut orci purus, interdum ornare nibh. Pellentesque pulvinar, nibh ac malesuada accumsan, urna nunc convallis tortor, ac vehicula nulla tellus eget nulla. Nullam lectus tortor, *consequat tempor hendrerit* quis, vestibulum in diam. Maecenas sed diam augue.
This statement requires citation [**?**].

Variational inference

Graphical Models

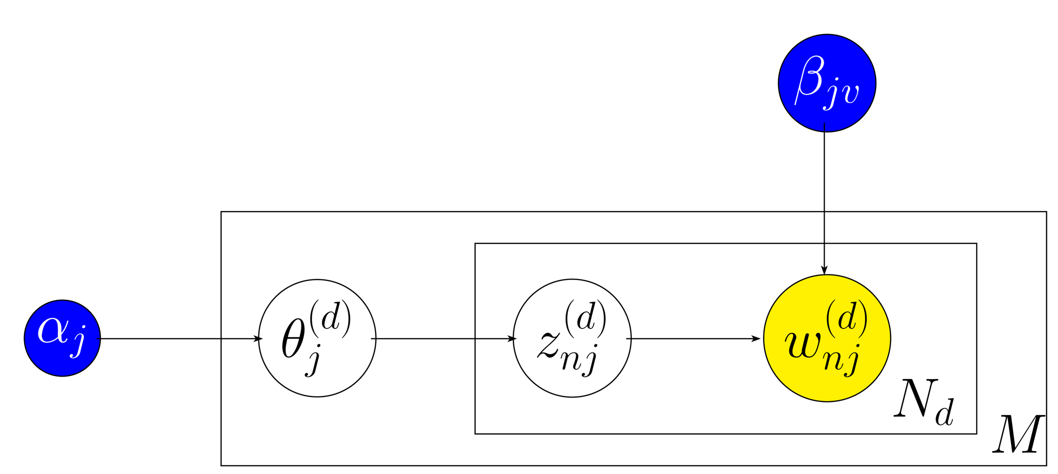

Figure 1: Generative model

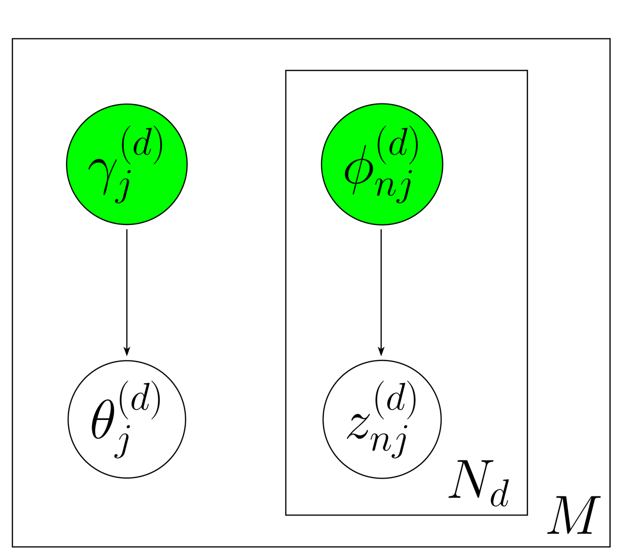

Figure 2: Variational model

E-step for a document d (Variational Inference Procedure)

- **Input**: a document d defined by its `word_incidences` $(w^{(d)})$, α, β
- **Output**: $\gamma^{(d)}$, $\phi^{(d)}$

Initialize $\phi_{ni}^{(d)} = \frac{1}{k}$ for all i and n.
Initialize $\gamma_i^{(d)} = \alpha + \frac{1}{k}\sum_{n=1}^{N_d} w_n^{(d)}$ for all i.
While *the expected log-likelihood for the document d has not converged*
  For $n = 1 \dots N_d$
    For $i = 1 \dots k$
      $\phi_{ni}^{(d)} = \beta_{iw_n^{(d)}} \exp(\Psi(\gamma_i^{(d)}))$
  Normalize $\phi_n^{(d)}$ to sum to 1.
  $\gamma^{(d)} = \alpha + \sum_{n=1}^{N_d} w_n^{(d)} \phi_n^{(d)}$

EM-algorithm

- **Input**: Corpus $\mathcal{D}$, number of topics k
- **Output**: α, β

For each $d \in \mathcal{D}$, compute $w^{(d)}$ (`word_incidences`).
Initialize α, β and $\Sigma_\gamma = 0$.
While *the expected log-likelihood has not converged*:
  For *each* $d \in \mathcal{D}$
    $(\gamma^{(d)}, \phi^{(d)}) = \textbf{E-step}(w^{(d)}, \alpha, \beta)$
    Update $\beta \leftarrow \beta + (\phi^{(d)})^\top w^{(d)}$
    Update $\Sigma_\gamma \leftarrow \Sigma_\gamma + \sum_{i=1}^k \Psi(\gamma_i^{(d)}) - \Psi\left(\sum_{j=1}^k \gamma_j^{(d)}\right)$
  Normalize β
  While α *has not converged*
    $\alpha \leftarrow \alpha - \frac{L'(\alpha)}{L''(\alpha)}$ where

$$\begin{cases} L'(\alpha) = |\mathcal{D}|k\left[\Psi(k\alpha) - \Psi(\alpha)\right] + \Sigma_\gamma \\ L''(\alpha) = |\mathcal{D}|k\left[k\Psi'(\alpha) - \Psi'(\alpha)\right] \end{cases}$$

Implementation issues

Results

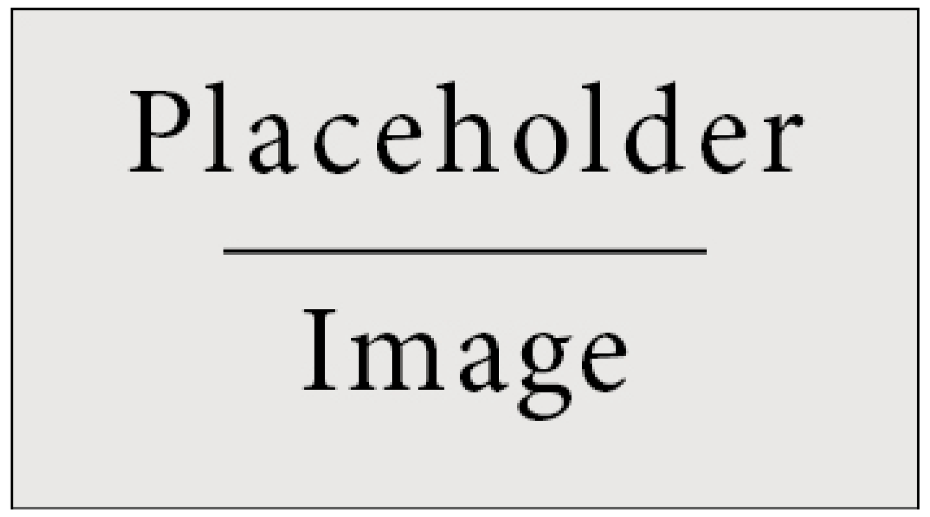

Figure 3: Figure caption

Nunc tempus venenatis facilisis. Curabitur suscipit consequat eros non porttitor. Sed a massa dolor, id ornare enim:

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 1: Table caption

Conclusion

Nunc tempus venenatis facilisis. **Curabitur suscipit** consequat eros non porttitor. Sed a massa dolor, id ornare enim. Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin*.

References

Contact Information

- `jerome{at}dockes.org`
- `pascal.lu{at}centraliens.net`