Project
Amith Ananthram (aa4461)
December 7, 2020

1 Motivation

2 Related Work

- LDA - Gaussian Embeddings - topic modeling for events

3 Modeling Events

noun and verb only, noun and verb and transliterated named entity, language news correction

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{x}) = \prod_{k}^{K} p(\mu_k) p(\Sigma_k) \prod_{i}^{N} (p(\theta_i) \prod_{j}^{M} (p(\zeta_{ij}|\theta_i) p(x_{ij}|\zeta_{ij}, \mu_{\zeta_{ij}}, \Sigma_{\zeta_{ij}}))$$

4 Inference

- Gibbs (Neal) - Variational Inference (only works when...) - Black-box VI: auto encoding variational Bayes, auto encoding variational Bayes for topic models

5 Experiments

5.1 Dataset

- talk about topic tags

5.2 Feature extraction

5.3 Hyperparameters

6 Results

derived complete conditionals, Gibbs sampler - implemented for model 1 (derived conditionals in Appendix), extremely slow - esp because operating per datapoint - auto encoding variational Bayes implementation

7 Conclusion

- future work

8 Appendix

8.1 Derivation of complete conditionals for μ_k and Σ_k

Below is the derivation of the complete conditional distributions for μ_k and Σ_k, the latent variables that govern our multivariate Gaussian distributions which model semantic sense (with Gaussian and inverse-Wishart priors, respectively). Note that these are the same for both variants 1 and 2 of our model as the bag of word embeddings is modeled independently from the bag of words.

W_k is the set of words across all documents assigned to topic k with size n_{W_k}. $\bar{w}_k = (1/n_{W_k}) \sum_l^{n_{W_k}} w_l$. Updated parameters for the conjugate complete conditionals are highlighted in green. Normalizing factors (that is, factors that are constant with respect to the random variable being modeled) which are dropped from the exponents (or in rare cases raised into the exponent) are highlighted in yellow.

$$
\begin{aligned}
p(\mu_k|\mu_{-k}, \boldsymbol{\Sigma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{x}) &\propto p(\mu_k) \prod_{w \in W_k} p(w|\mu_k, \Sigma_k) \\
&\propto \exp -\frac{1}{2}(\mu_k - m_k)^T S_k^{-1} (\mu_k - m_k) \prod_{w \in W_k} \exp -\frac{1}{2}(w - \mu_k)^T \Sigma_k^{-1} (w - \mu_k) \\
&\propto \exp -\frac{1}{2}(\mu_k - m_k)^T S_k^{-1} (\mu_k - m_k) \exp \sum_{w \in W_k} -\frac{1}{2}(w - \mu_k)^T \Sigma_k^{-1} (w - \mu_k) \\
&\propto \exp -\frac{1}{2}((\mu_k - m_k)^T S_k^{-1} (\mu_k - m_k) + \sum_{w \in W_k} (w - \mu_k)^T \Sigma_k^{-1} (w - \mu_k)) \\
&\propto \exp -\frac{1}{2}((\mu_k - m_k)^T S_k^{-1} (\mu_k - m_k) + \sum_{w \in W_k} (w^T \Sigma_k^{-1} w - 2\mu_k^T \Sigma_k^{-1} w + \mu_k^T \Sigma^{-1} \mu_k)) \\
&\propto \exp -\frac{1}{2}((\mu_k - m_k)^T S_k^{-1} (\mu_k - m_k) - 2\mu_k^T \Sigma_k^{-1} n_{W_k} \bar{w}_k + n_{W_k} \mu_k^T \Sigma^{-1} \mu_k) \\
&\propto \exp -\frac{1}{2}(\mu_k^T (S_k^{-1} + n_{W_k} \Sigma^{-1}) \mu_k - 2\mu_k^T (S_k^{-1} m_k + \Sigma_k^{-1} n_{W_k} \bar{w}_k) + m_k^T S_k^{-1} m_k) \\
&\propto \exp -\frac{1}{2}(\mu_k^T \hat{S}_k^{-1} \mu_k - 2\mu_k^T \hat{S}_k^{-1} \hat{S}_k (S_k^{-1} m_k + \Sigma_k^{-1} n_{W_k} \bar{w}_k)) \\
&\propto \exp -\frac{1}{2}(\mu_k^T \hat{S}_k^{-1} \mu_k - 2\mu_k^T \hat{S}_k^{-1} \hat{m}_k + \hat{m}_k^T \hat{S}_k^{-1} \hat{m}_k) \\
&= \mathcal{N}(\hat{m}_k, \hat{S}_k), \hat{m}_k = \hat{S}_k (S_k^{-1} m_k + \Sigma_k^{-1} n_{W_k} \bar{w}_k), \hat{S}_k^{-1} = S_k^{-1} + n_{W_k} \Sigma_k^{-1}
\end{aligned}
$$

$$
\begin{aligned}
p(\Sigma_k|\boldsymbol{\mu}, \Sigma_{-k}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{x}) &\propto p(\Sigma_k) \prod_{w \in W_k} p(w|\mu_k, \Sigma_k) \\
&\propto |\Sigma_k|^{-(\nu_k+d+1)/2} \exp -\frac{1}{2} \operatorname{Tr}(\Psi_k \Sigma_k^{-1}) \prod_{w \in W_k} |\Sigma_k|^{-1/2} \exp -\frac{1}{2}(w-\mu_k)^T \Sigma_k^{-1}(w-\mu_k) \\
&\propto |\Sigma_k|^{-(n_{W_k}+\nu_k+d+1)/2} \exp -\frac{1}{2}(\operatorname{Tr}(\Psi_k \Sigma_k^{-1}) + \sum_{w \in W_k} (w-\mu_k)^T \Sigma_k^{-1}(w-\mu_k)) \\
&\propto |\Sigma_k|^{-(n_{W_k}+\nu_k+d+1)/2} \exp -\frac{1}{2}(\operatorname{Tr}(\Psi_k \Sigma_k^{-1}) + \sum_{w \in W_k} \operatorname{Tr}((w-\mu_k)^T \Sigma_k^{-1}(w-\mu_k))) \\
&\propto |\Sigma_k|^{-(n_{W_k}+\nu_k+d+1)/2} \exp -\frac{1}{2}(\operatorname{Tr}(\Psi_k \Sigma_k^{-1}) + \sum_{w \in W_k} \operatorname{Tr}((w-\mu_k)(w-\mu_k)^T \Sigma_k^{-1})) \\
&\propto |\Sigma_k|^{-(n_{W_k}+\nu_k+d+1)/2} \exp -\frac{1}{2}(\operatorname{Tr}(\Psi_k \Sigma_k^{-1} + \sum_{w \in W_k} (w-\mu_k)(w-\mu_k)^T \Sigma_k^{-1})) \\
&\propto |\Sigma_k|^{-(n_{W_k}+\nu_k+d+1)/2} \exp -\frac{1}{2}(\operatorname{Tr}((\Psi_k + \sum_{w \in W_k} (w-\mu_k)(w-\mu_k)^T)\Sigma_k^{-1})) \\
&= \mathscr{W}^{-1}(\hat{\Psi}_k, \hat{\nu}_k), \hat{\Psi}_k = \Psi_k + \sum_{w \in W_k} (w-\mu_k)(w-\mu_k)^T, \hat{\nu}_k = n_{W_k} + \nu_k
\end{aligned}
$$