CS247: Advanced Data Mining (Winter 2024) January 24, 2024

Assignment 2 (Part 2)

Due Date: **January 31, 2024**

Instructions

- Submit your answer on Gradescope as a PDF file. Both typed and scanned handwritten answers are acceptable.
- Submit your solutions to Part 1 and Part 2 through GradeScope in BruinLearn separately.
- Late submissions are allowed up to 24 hours post-deadline with a penalty factor of $\mathbf{1}(t \leq 24)e^{-(\ln(2)/12)t}$.
- Ensure all sources are cited appropriately; plagiarism will be reported.

Problems

Problem 1: Statistics Review (20 points)

1. **(10 points)** Recall the Dirichlet distribution, a family of continuous multivariate probability distribution used as prior to the multinomial distribution, with the PDF specified as:

$$f(\mathbf{x}; \alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_{i-1}},$$

   where $\Gamma(x)$ is the Gamma function with derivative $\Gamma(x)' = \phi(x)$, the digamma function. Derive the maximum likelihood estimator for parameter α for the Dirichlet distribution.

2. **(10 points)** Use the *pdf kernel method* to derive the posterior distribution the following conjugate pair: prior $p(\theta) = \mathcal{N}(\theta; 0, I)$ and likelihood function $p(x; \theta) = \mathcal{N}(x; \mu, \Sigma)$. Show all steps.

Answer:

Problem 2: Topic Models (30 points)

1. **(5 points)** Discuss the similarities and differences between the multinomial mixture model and the mixture of unigrams model.

2. **(15 points)** Derive the following quantities in the pLSA model and show all your steps.

   (a) The joint density of all random variables $p(w, d, z; \theta, \beta)$.

   (b) The conditional density $p(z \mid w, d; \theta, \beta)$.

   (c) The conditional density of $p(w \mid z, d; \theta, \beta)$.

   Note: in this notation, θ, β are given as model hyperparameters, hence they are separated from the random variable using ';'.

3. **(10 points)** Write the log-likelihood function of the pLSA model as sum of two terms: the Evidence Lower Bound (ELBO) and the KL divergence between a variational distribution and the latent posterior. Point out each term in your answer.

Answer: