Homework 3 Part 1

This is an individual assignment.

Solve all problems by hand. You may type your answers in markdown cells or `push` a PDF file with your handwritten answers.

Problem 1 (20 points)

Consider a model in which the set of all hidden variables are denoted by $\mathbf{z}$.

Let x be a sample observation (single sample, e.g., $x = 1.5$). The data likelihood is defined as

$$p(x|z) \sim \mathcal{N}\left(x|\mu, \sigma^2 = 1\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-z)^2}{2}\right)$$

and the prior probability on the latent variables is defined as an Exponential with parameter $\lambda = 1$,

$$p(z|\lambda) \sim \mathrm{Exp}(\lambda = 1) = \exp(-z), \quad z \geq 0$$

We saw that the resulting posterior probability, $p(z|x)$, is *intractable*. Suppose we use variational inference with the following proposed family of surrogate posteriors

$$q(z) \sim \mathrm{Exp}(\alpha) = \alpha \exp(-\alpha z), \quad z \geq 0.$$

That is, for a specific value of α, $q(z)$ represents a proposed surrogate posterior.

Suppose you only have 1 data sample $x = 1.5$, find the value for α that maximizes the ELBO for this surrogate posterior.

$$\mathbb{E}_{z\sim q(z)}[z] = \frac{1}{\alpha}, \quad \text{where} \quad z \sim e^{\alpha}$$

$$\mathbb{E}_{z}[z^2] = \frac{2}{\alpha^2}$$

$$\mathbb{E}_{z}[a \cdot z] = a \cdot \mathbb{E}_{z}[z] = a \cdot \frac{1}{\alpha}$$

$$\mathbb{E}_{z}[a] = a$$

HW3 - P1 - Q1

Evidence lower bound (ELBO)

$$\mathcal{L} = \mathop{\mathbb{E}}_{z \sim q(z)}\left[\ln p(z|x) - \ln q(z)\right] = \mathop{\mathbb{E}}_{z}\left[\ln\left(\frac{p(x|z) \cdot p(z)}{p(x)}\right)\right]$$

$q(z)$ = surrogate posterior

$p(z|x)$ = true intractable posterior

$$- \mathop{\mathbb{E}}_{z}\left[\ln q(z)\right]$$

$$= \mathop{\mathbb{E}}_{z}\left[\ln p(x|z)p(z)\right] - \mathop{\mathbb{E}}_{z}\left[\ln(p(x))\right] - \mathop{\mathbb{E}}_{z}\left[\ln q(z)\right]$$

① ... $\ln(p(x|z)\,p(z)) = \ln\left[\frac{1}{\sqrt{2\pi}} e^{-\left(\frac{(x-z)^2}{2}\right)} \cdot e^{-z}\right]$

$$= \ln\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-z)^2}{2} - z}\right]$$

$$= \ln\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{\left(-x^2 + 2xz - z^2 - 2z\right)}{2}$$

$$\ln(p(x|z)p(z)) = \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{x^2}{2} + (x-1)z - \frac{z^2}{2}$$

---

② ... $\ln q(z) = \ln\left[\alpha e^{-\alpha z}\right] = \ln \alpha - \alpha z$

$$\mathcal{L} = \mathbb{E}_z\left[\ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{x^2}{2} + (x-1)z - \frac{z^2}{2}\right] - \ln(p(x))$$
$$- \mathbb{E}_z\left[\ln \alpha - \alpha z\right]$$

$$= \ln\frac{1}{\sqrt{2\pi}} - \frac{x^2}{2} + \mathbb{E}_z\left[(x-1)z\right] - \mathbb{E}_z\left[\frac{z^2}{2}\right] - \ln(p(x))$$
$$- \ln \alpha + \mathbb{E}_z\left[\alpha z\right]$$

$$= \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \ln\alpha - \frac{x^2}{2} + (x-1)\frac{1}{\alpha} - \frac{1}{2}\left(\frac{2}{\alpha^2}\right) - \ln p(x)$$
$$- \ln\alpha + \cancel{\alpha}\frac{1}{\cancel{\alpha}}$$

$\underline{x = 1.5}$

$$\mathcal{L} = \ln\frac{1}{\sqrt{2\pi}} - 2\ln\alpha - \frac{(1.5)^2}{2} + \frac{0.5}{\alpha} - \frac{1}{\alpha^2} - \ln p(x) + 1$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \Rightarrow 0 - 2\frac{1}{\alpha} - 0 + \frac{0.5(-1)}{\alpha^2} + \frac{2}{\alpha^3} = 0$$

$$\Rightarrow -2\alpha^2 - 0.5\alpha + 2 = 0$$
$$\Rightarrow \alpha^2 + 0.5\alpha - 2 = 0$$
$$\Rightarrow 2\alpha^2 + \alpha - 4 = 0$$
$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-1 \pm \sqrt{1 + 4(2)4}}{2(2)} = \frac{-1 \pm \sqrt{33}}{4}$$

$$\boxed{\alpha = \frac{-1+\sqrt{33}}{4}, \frac{-1-\sqrt{33}}{4}}$$

$$= 1.186, -1.686$$

Problem 2 (25 points)

Consider a model in which the set of all hidden stochastic variables, denoted collectively by $\mathbf{Z}$, comprises some latent variables $\mathbf{z}$ together with some model parameters θ.

Suppose we use a variational distribution that factorizes between latent variables and parameters so that

$$q(\mathbf{z}, \theta) = q_{\mathbf{z}}(\mathbf{z}) q_\theta(\theta),$$

in which the distribution $q_\theta(\theta)$ is approximated by a point estimate of the form $q_\theta(\theta) = \delta(\theta - \theta_0)$ where θ_0 is a vector of free parameters.

Show that variational optimization of this factorized distribution is equivalent to an EM algorithm, in which the E-step optimizes $q_{\mathbf{z}}(\mathbf{z})$, and the M-step maximizes the expected complete-data log posterior distribution of θ with respect to θ_0.

Q2. $q(z,\theta) = q_z(z)\, q_\theta(\theta)$, where $q_\theta(\theta) = \delta(\theta - \theta_0)$ vector of free params

$P(X,Z) = \underbrace{P(X|Z)}_{\text{likelihood}} P(Z)$ ← prior

EM maximizes the log likelihood of the observed data w.r.t. Z and model parameter θ.

Factorizing variational distribution between Z and θ.

$q(z,\theta) = q(z)\, \underbrace{q_\theta(\theta)}$

approx. by point estimate $q_\theta = \delta(\theta - \theta_0)$
here θ_0 vector free parameter.

E step: compute the posterior dist. of latent variables given observed data and current estimates of model parameters.

$\therefore\ q(z|X,\theta_0) = q_z(z)$

$\therefore$ E step optimizes the varicational distribution $q_z(z)$ by computing the posterior distribution of the latent variables with observed data and current estimates of the model parameters:

$$q_z(z) \propto p(X, z : \theta_0)$$

This is equivalent of E-step in EM algorithm.

In M-step, we find parameters that maximize the expected complete data log posterior distribution.

$\theta_0 \leftarrow \arg\max_\theta E[\log P(X, z : \theta) | X, \theta_0)]$
where expectation is taken w.r.t. variational distribution $q(z, \theta)$

$$\therefore \log p(X, z : \theta) = \log p(X|z) + \log P(z)$$

$$\log(X|z) = \underbrace{\sum_i \log p(X_i | z ; \theta)}_{\text{summing over all observations in the data set}}$$

Plugging in the factorized variational distribution and taking the expectation w.r.t. $q_2(z)$

$\cancel{E[\log p(X|z)] = \sum_i \int \log p(x_i|z,\theta)\, q}$

$q_2(z) \propto p(X, z : \theta_0)$

$\theta_0 \leftarrow \arg\max_\theta \int \log P(X, z : \theta)\, q_2(z)\, dz$

simplifying the integral :

$\theta_0 \leftarrow \arg\max_\theta \log P(X|z : \theta)\, q_2(z) =$

$= \arg\max_\theta \sum_i \log (x_i, z : \theta)\, q_2(z)$

This is equivalent of M-step.

$\therefore$ Variational optimization of factorized distribution is equivalent of EM algorithm

---

On-Time (5 points)

Submit your assignment before the deadline.

Submit Your Solution

Confirm that you've successfully completed the assignment.

Along with the Notebook, include a PDF of the notebook with your solutions.

`add` and `commit` the final version of your work, and `push` your code to your GitHub repository.

Submit the URL of your GitHub Repository as your assignment submission on Canvas.
