Homework 3 Part 1 - Solutions

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

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

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

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

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

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

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.

The *Evidence Lowe Bound (ELBO)* is given by:

$$\begin{aligned}
\mathcal{L} &= \mathbb{E}_{z\sim q(z)}[\ln p(z|x) - \ln q(z)] \\
&= \mathbb{E}_{z\sim q(z)}[lnp(z|x)] - \mathbb{E}_{z\sim q(z)}[\ln q(z)] \\
&= \mathbb{E}_{z\sim q(z)}\left[\ln\left(\frac{p(x|z)p(z)}{p(x)}\right)\right] - \mathbb{E}_{z\sim q(z)}[\ln q(z)] \\
&= \mathbb{E}_{z\sim q(z)}[\ln p(x|z)] + \mathbb{E}_{z\sim q(z)}[\ln p(z)] - \mathbb{E}_{z\sim q(z)}[\ln p(x)] - \mathbb{E}_{z\sim q(z)}[q(z)] \\
&= \mathbb{E}_{z\sim q(z)}[\ln p(x|z)] + \mathbb{E}_{z\sim q(z)}[\ln p(z)] - \ln p(x) - \mathbb{E}_{z\sim q(z)}[q(z)] \\
&= \mathbb{E}_{z\sim q(z)}[\ln p(x|z)] + \mathbb{E}_{z\sim q(z)}[\ln p(z)] - \mathbb{E}_{z\sim q(z)}[q(z)] + \text{const.}
\end{aligned}$$

since $p(x)$ is fixed given a dataset and it does not depend on α, we treat it as a constant.

where the data likelihood is given as:

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

the prior on the latent variables,

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

and the surrogate posterior is

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

Moreover, we are given an observation/sample $x_1 = 1.5$.

Substituting,

$$\begin{aligned}
\mathcal{L} &= \mathbb{E}_{z\sim q(z)}[\ln p(x|z)] + \mathbb{E}_{z\sim q(z)}[\ln p(z)] - \ln p(x) - \mathbb{E}_{z\sim q(z)}[q(z)] \\
&= \mathbb{E}_{z\sim q(z)}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}(x-z)^2\right] + \mathbb{E}_{z\sim q(z)}[-z] - \mathbb{E}_{z\sim q(z)}[\ln\alpha - \alpha z] \\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\mathbb{E}_{z\sim q(z)}[x^2 - 2xz + z^2] - \mathbb{E}_{z\sim q(z)}[z] - \ln\alpha + \alpha\mathbb{E}_{z\sim q(z)}[z] \\
&= -\frac{1}{2}\ln(2\pi) - \frac{x^2}{2} + x\mathbb{E}_{z\sim q(z)}[z] - \frac{1}{2}\mathbb{E}_{z\sim q(z)}[z^2] - \mathbb{E}_{z\sim q(z)}[z] - \ln\alpha + \alpha\mathbb{E}_{z\sim q(z)}[z] \\
&= -\frac{1}{2}\ln(2\pi) - \ln\alpha - \frac{x^2}{2} + (x - 1 + \alpha)\mathbb{E}_{z\sim q(z)}[z] - \frac{1}{2}\mathbb{E}_{z\sim q(z)}[z^2]
\end{aligned}$$

Since $z \sim q(z)$ and $q(z) \sim \mathrm{Exp}(\alpha)$, $z \geq 0$, we know that

$$\begin{aligned}
\mathbb{E}_{z\sim q(z)}[z] &= \frac{1}{\alpha} \\
\mathbb{E}_{z\sim q(z)}[z^2] &= \frac{2}{\alpha^2}
\end{aligned}$$

Substituting,

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{2}\ln(2\pi) - \ln\alpha - \frac{x^2}{2} + (x - 1 + \alpha)\mathbb{E}_{z\sim q(z)}[z] - \frac{1}{2}\mathbb{E}_{z\sim q(z)}[z^2] \\
&= -\frac{1}{2}\ln(2\pi) - \ln\alpha - \frac{x^2}{2} + (x - 1 + \alpha)\frac{1}{\alpha} - \frac{1}{2}\frac{2}{\alpha^2} \\
&= -\frac{1}{2}\ln(2\pi) - \ln\alpha - \frac{x^2}{2} + (x - 1)\frac{1}{\alpha} + 1 - \frac{1}{\alpha^2}
\end{aligned}$$

Optimizing for $alpha$:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \alpha} &= 0 \\
\iff -\frac{1}{\alpha} - \frac{x-1}{\alpha^2} + \frac{2}{\alpha^3} &= 0 \\
\iff -\alpha^2 + 2 - \alpha(x-1) &= 0 \\
\iff \alpha^2 + \alpha(x-1) - 2 &= 0 \\
\iff \alpha = -\frac{x-1}{2} \pm \sqrt{\frac{(x-1)^2+8}{2}}
\end{aligned}
$$

Since $\alpha \geq 0$ then we have: $\alpha^* = -\frac{x-1}{2} + \sqrt{\frac{(x-1)^2+8}{2}}$.

Given observed data $x = 1.5$, we find:

$$
\alpha^* = -\frac{1.5-1}{2} + \sqrt{\frac{(1.5-1)^2+8}{2}} \approx 1.781
$$

In [3]:
```
import numpy as np

-(1.5-1)/2 + np.sqrt(((1.5-1)**2+8)/2)
```

Out[3]:
```
1.7810096011589902
```

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.

(From Bishop, C. (2006) Pattern Recognition and Machine Learning, Springer. Exercise 10.5.)

Recall that the optimal ELBO occurs when the KL-divergence between the posterior $p(z|x)$ and the surrogate posterior $q(z)$ is minimized. The minimum (0) occurs when $p(Z|x) \approx q(z)$.

The KL-divergence between $p(Z|x)$ and $q(Z)$ is given by:

$$\begin{aligned} \mathrm{KL}\left(q(Z) \parallel p(Z|x)\right) &= \int_Z q(Z) \ln\left(\frac{q(Z)}{p(Z|x)}\right) dZ \\ \mathrm{KL}\left(q \parallel p\right) &= \int_Z q(Z) \ln\left(\frac{q(Z)}{p(Z|x)}\right) dZ \end{aligned}$$

where $Z = \{\mathbf{z}, \theta\}$ and

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

Substituting,

$$\begin{aligned} \mathrm{KL}\left(q \parallel p\right) &= \int_{\mathbf{z}} \int_\theta q(\mathbf{z}, \theta) \ln\left(\frac{q(\mathbf{z}, \theta)}{p(\mathbf{z}, \theta|x)}\right) d\mathbf{z} d\theta \\ &= -\int_{\mathbf{z}} \int_\theta q(\mathbf{z}, \theta) \ln\left(\frac{p(\mathbf{z}, \theta|x)}{q(\mathbf{z}, \theta)}\right) d\mathbf{z} d\theta \\ &= -\int_{\mathbf{z}} \int_\theta q(\mathbf{z}) q(\theta) \ln\left(\frac{p(\mathbf{z}, \theta|x)}{q(\mathbf{z}) q(\theta)}\right) d\mathbf{z} d\theta \end{aligned}$$

We are also given that

$$q(\theta) = \delta(\theta - \theta_0) = \begin{cases} 1, & \theta = \theta_0 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} P(\theta = \theta_0) = 1 \\ P(\theta \neq \theta_0) = 0 \end{cases}$$

Thus, KL-divergence will be non-zero if $\theta = \theta_0$, i.e., θ_0 is a random instantiation of the random variable θ.

$$\begin{aligned} \mathrm{KL}\left(q \parallel p\right) &= -\int_{\mathbf{z}} \int_\theta q(\mathbf{z}) q(\theta) \ln\left(\frac{p(\mathbf{z}, \theta|x)}{q(\mathbf{z}) q(\theta)}\right) d\mathbf{z} d\theta \\ &= -\int_{\mathbf{z}} q(\mathbf{z}) \ln\left(\frac{p(\mathbf{z}, \theta_0|x)}{q(\mathbf{z})}\right) d\mathbf{z} \\ &= -\int_{\mathbf{z}} q(\mathbf{z}) \ln\left(\frac{p(\mathbf{z}|\theta_0, x) p(\theta_0|x)}{q(\mathbf{z}) p(x)}\right) d\mathbf{z} \\ &= -\int_{\mathbf{z}} q(\mathbf{z}) \ln\left(\frac{p(\mathbf{z}|\theta_0, x) p(\theta_0|x)}{q(\mathbf{z})}\right) d\mathbf{z} + \text{const.} \\ &= -\int_{\mathbf{z}} q(\mathbf{z}) \ln\left(\frac{p(\mathbf{z}|\theta_0, x)}{q(\mathbf{z})}\right) d\mathbf{z} + \text{const.} \end{aligned}$$

We see that this KL-divergence is minimized when $q(\mathbf{z}) = p(\mathbf{z}|\theta_0, x)$. This is the **E-step in the EM algorithm** where we fix parameter θ_0 and optimize the (surrogate) posterior probability on the latent variable $\mathbf{z}$.

To find $q(\theta)$, we use the solution derived in lecture 17:

$$\begin{aligned}
&\int_\theta q(\theta) \int_\mathbf{z} q(\mathbf{z}) \ln\left(\frac{p(x,\theta,\mathbf{z}}{q(\theta)q(\mathbf{z})}\right) d\mathbf{z} d\theta \\
&= \int_\theta q(\theta) \mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}[\ln p(x,\theta,\mathbf{z})] d\theta - \int_\theta q(\theta) \ln q(\theta) d\theta + \text{const}.
\end{aligned}$$

The optimization problem is then reduced to maximizing expected complete log posterior distribution

$$\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z})}[\ln p(x,\theta,\mathbf{z})],$$

with respect to θ_0, which is equivalent to the **M-step of the 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.
