Homework 5

Leroy Souz

December 16, 2024

1 Problem 1

1.1

We can calculate the log-likehood ℓ by using the formula:

$$\ell = log(\sum_{k=0}^{K-1} \pi_k \mathcal{N}(x_n|\mu_k, \sigma_k^2)) \tag{1}$$

We have 4 data points $x_n = \{1, 2, 3, 4\}$ and 2 clusters $K = 2$. Thus:
n = 1

$$\begin{aligned} \ell_1 &= log(\pi_1 \mathcal{N}(1|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(1|\mu_2, \sigma_2^2)) \\ &= log(0.5\mathcal{N}(2|1, 1) + 0.5\mathcal{N}(2| - 1, 1)) \\ &= -2.09 \end{aligned}$$

n = 2

$$\begin{aligned} \ell_2 &= log(\pi_1 \mathcal{N}(2|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(2|\mu_2, \sigma_2^2)) \\ &= log(0.5\mathcal{N}(1|1, 1) + 0.5\mathcal{N}(1| - 1, 1)) \\ &= -1.48 \end{aligned}$$

n = 3

$$\begin{aligned} \ell_3 &= log(\pi_1 \mathcal{N}(3|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(3|\mu_2, \sigma_2^2)) \\ &= log(0.5\mathcal{N}(-1|1, 1) + 0.5\mathcal{N}(-1| - 1, 1)) \\ &= -1.48 \end{aligned}$$

n = 4

$$\begin{aligned} \ell_4 &= log(\pi_1 \mathcal{N}(4|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(4|\mu_2, \sigma_2^2)) \\ &= log(0.5\mathcal{N}(-2|1, 1) + 0.5\mathcal{N}(-2| - 1, 1)) \\ &= -2.09 \end{aligned}$$

Adding these all up we get $\ell = -7.158$

1.2

To compute γ_{n0} and γ_{n1} we can use the formula:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \sigma_k^2)}{\sum_{j=0}^{K-1} \pi_j \mathcal{N}(x_i|\mu_j, \sigma_j^2)}$$

n = 1

$$\begin{aligned} \gamma_{10} &= \frac{0.5\mathcal{N}(2|1, 1)}{0.5\mathcal{N}(2|1, 1) + 0.5\mathcal{N}(2| - 1, 1)} \\ &= 0.984 \end{aligned}$$

$$\gamma_{11} = \frac{0.5\mathcal{N}(2|-1,1)}{0.5\mathcal{N}(2|1,1) + 0.5\mathcal{N}(2|-1,1)} = 0.018$$

n = 2

$$\gamma_{20} = \frac{0.5\mathcal{N}(1|1,1)}{0.5\mathcal{N}(1|1,1) + 0.5\mathcal{N}(1|-1,1)} = 0.881$$

$$\gamma_{21} = \frac{0.5\mathcal{N}(1|-1,1)}{0.5\mathcal{N}(1|1,1) + 0.5\mathcal{N}(1|-1,1)} = 0.119$$

n = 3

$$\gamma_{30} = \frac{0.5\mathcal{N}(-1|1,1)}{0.5\mathcal{N}(-1|1,1) + 0.5\mathcal{N}(-1|-1,1)} = 0.881$$

$$\gamma_{31} = \frac{0.5\mathcal{N}(-1|-1,1)}{0.5\mathcal{N}(-1|1,1) + 0.5\mathcal{N}(-1|-1,1)} = 0.119$$

n = 4

$$\gamma_{40} = \frac{0.5\mathcal{N}(-2|1,1)}{0.5\mathcal{N}(-2|1,1) + 0.5\mathcal{N}(-2|-1,1)} = 0.984$$

$$\gamma_{41} = \frac{0.5\mathcal{N}(-2|-1,1)}{0.5\mathcal{N}(-2|1,1) + 0.5\mathcal{N}(-2|-1,1)} = 0.018$$

1.3

To compute the new $\pi_0(t+1)$, $\pi_1(t+1)$, $\mu_0(t+1)$, $\mu_1(t+1)$, $\sigma_0^2(t+1)$, and $\sigma_1^2(t+1)$ we can use the formulas:

$$\pi_k(t+1) = \frac{\sum_{n=1}^N \gamma_{nk}}{N}$$
$$\mu_k(t+1) = \frac{\sum_{n=1}^N \gamma_{nk} x_n}{\sum_{n=1}^N \gamma_{nk}}$$
$$\sigma_k^2(t+1) = \frac{\sum_{n=1}^N \gamma_{nk}(x_n - \mu_k)^2}{\sum_{n=1}^N \gamma_{nk}}$$

New $\pi_0(t+1)$

$$\pi_0(t+1) = \frac{\sum_{n=1}^N \gamma_{n0}}{N} = \frac{2}{4} = 0.5$$

New $\pi_1(t+1)$

$$\pi_1(t+1) = \frac{\sum_{n=1}^N \gamma_{n1}}{N} = \frac{2}{4} = 0.5$$

New $\mu_0(t+1)$

$$\begin{aligned}\mu_0(t+1) &= \frac{\sum_{n=1}^{N} \gamma_{n0} x_n}{\sum_{n=1}^{N} \gamma_{n0}} \\ &= \frac{0.018*2+0.119*1+0.883*-1+0.984*-2}{2.004} \\ &= -1.34\end{aligned}$$

New $\mu_1(t+1)$

$$\begin{aligned}\mu_1(t+1) &= \frac{\sum_{n=1}^{N} \gamma_{n1} x_n}{\sum_{n=1}^{N} \gamma_{n1}} \\ &= 1.34\end{aligned}$$

New $\sigma_0^2(t+1) = 0.69$
New $\sigma_1^2(t+1) = 0.69$

1.4

After using these new parameters, we get the log likehood $\ell = -6.463$. We can see that the log-likehood has increased from -7.158 to -6.463.

2 Problem 2

2.1

To calculate $P(C|S=T, W=T)$, using VE, we eliminate R.
For C = T:

$$\begin{aligned}R_1 &= \sum_R P(R|C=T)P(W=T|S=T,R) \\ &= P(R=T|C=T)P(W=T|S=T,R=T) + P(R=F|C=T)P(W=T|S=T,R=F) \\ &= 0.8*0.99+0.2*0.9 \\ &= 0.972\end{aligned}$$

For C = F:

$$\begin{aligned}R_2 &= \sum_R P(R|C=F)P(W=T|S=T,R) \\ &= P(R=T|C=F)P(W=T|S=T,R=T) + P(R=F|C=F)P(W=T|S=T,R=F) \\ &= 0.2*0.99+0.8*0.9 \\ &= 0.918\end{aligned}$$

Combining $P(C)$ and $P(S=T|C)$: For C = T:

$$\begin{aligned}&= P(C=T)P(S=T|C=T)R_1 \\ &= 0.5*0.1*0.972 \\ &= 0.0486\end{aligned}$$

For C = F:

$$\begin{aligned}&= P(C=F)P(S=T|C=F)R_2 \\ &= 0.5*0.5*0.918 \\ &= 0.2295\end{aligned}$$

We now normalize these values to get the final probability:

$$P(C=T|S=T,W=T) = 0.175, P(C=F|S=T,W=T) = 0.825$$

2.2

Through Gibbs Sampling, the following posteriors are calculated::

$$P(C = T|S = T, W = T) = 0.172, P(C = F|S = T, W = T) = 0.824$$

3 Problem 3

Let's start with the decomposition of the marginal log-likelihood $\log p_\theta(x_i)$:

$$\log p_\theta(x_i) = \log \sum_z p_\theta(x_i, z)$$

We multiply and divide by a variational distribution $q_\phi(z|x_i)$ (any non-negative function such that $\sum_z q_\phi(z|x_i) = 1$):

$$\log p_\theta(x_i) = \log \sum_z \frac{p_\theta(x_i, z)q_\phi(z|x_i)}{q_\phi(z|x_i)}$$

Using Jensen's inequality, we can write:

$$\log p_\theta(x_i) \geq \sum_z q_\phi(z|x_i) \log \frac{p_\theta(x_i, z)}{q_\phi(z|x_i)}$$

By simplifying the fraction inside the logarithm we get:

$$\mathbb{E}_{q_\phi(z|x_i)} \left[\log \frac{p_\theta(x_i, z)}{q_\phi(z|x_i)} \right] = \mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(x_i, z) - \log q_\phi(z|x_i) \right]$$

The joint distribution $p_\theta(x_i, z)$ can be rewritten as $p_\theta(x_i, z) = p_\theta(z)p_\theta(x_i|z)$. Substituting this:

$$\mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(x_i, z) - \log q_\phi(z|x_i) \right] = \mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(z) + \log p_\theta(x_i|z) - \log q_\phi(z|x_i) \right]$$

Separate terms under the expectation:

$$\mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(z) - \log q_\phi(z|x_i) \right] + \mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(x_i|z) \right]$$

The first term corresponds to the Kullback-Leibler (KL) divergence between $q_\phi(z|x_i)$ and $p_\theta(z)$:

$$\mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(z) - \log q_\phi(z|x_i) \right] = -D_{\mathrm{KL}}(q_\phi(z|x_i) \| p_\theta(z))$$

The second term remains as the conditional expectation:

$$\mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(x_i|z) \right]$$

Combining these, we have the Evidence Lower Bound (ELBO):

$$\log p_\theta(x_i) \geq -D_{\mathrm{KL}}(q_\phi(z|x_i) \| p_\theta(z)) + \mathbb{E}_{q_\phi(z|x_i)} \left[\log p_\theta(x_i|z) \right]$$

4 Problem 4

4.1

Bipolar coding allows a way to mitigate the positive bias found in recommendation systems data by centering the scale around 0.

4.2

To modify the formulations from binary to bipolar codings, we can replace the sigmoid functions to a tanh function. Thus the functions modify to:

$$P(h_1 = 1|m_1, m_2, m_3) = \tanh \frac{(m^t W[:, 1] + c_1)}{2}$$

$$P(m_1 = 1|h_1, m_2, m_3) = \tanh \frac{(W[1, :]h + b_1)}{2}$$

The 1/2 scaling factor is used to ensure that the output of the tanh function is between -1 and 1.

4.3

Based on user prefernece he will dislike the movie