Machine Learning 2 — Homework 6

Maurice Frank
11650656
maurice.frank@posteo.de

October 12, 2019

Problem 1.

We have PDF $p(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$ and the approximation of it $q(\boldsymbol{x})$.

a)

Algorithm 1 Rejection Sampler

function SAMPLE_REJECTION(q, $\tilde{p}$, c)
  $acc \leftarrow False$
  while $acc == False$ **do**
    $z_0 \leftarrow \text{sample}(q(z))$
    $u \leftarrow \text{rand}[0, c \cdot q(z_0)]$
    $acc \leftarrow u \leq \tilde{p}(z)$
  end while
  return z
end function

b)

Yes the generated samples are independent. We sample independently from $q(\boldsymbol{z})$ which makes the accepted samples also independently.

c)

$$w_n = \frac{p(z_n)/q(z_n)}{\sum_m p(z_m)/q(z_m)}$$

d)

$$
\begin{aligned}
\alpha(x_{t+1}, x_t) &= \min\left(1, \frac{\tilde{p}(x_{t+1})q(x_t|x_{t+1})}{\tilde{p}(x_t)q(x_{t+1}|x_t)}\right) \\
&= \min\left(1, \frac{\tilde{p}(x_{t+1})q(x_t|x_{t+1})}{\tilde{p}(x_t)q(x_{t+1})}\right) \\
&= \min\left(1, \frac{\tilde{p}(x_{t+1})q(x_{t+1}|x_t)q(x_t)}{\tilde{p}(x_t)q(x_{t+1})^2}\right) \\
&= \min\left(1, \frac{\tilde{p}(x_{t+1})q(x_t)}{\tilde{p}(x_t)q(x_{t+1})}\right) \\
&= \min\left(1, \frac{p(x_{t+1})q(x_t)}{p(x_t)q(x_{t+1})}\right)
\end{aligned}
$$

e)

A proposed sample using the proposal distribution is chosing independent of the previous sample: $q(x_{t+1}|x_t) = q(x_t)$. Whether or not the proposal is added to the chain though, is dependent on the previous sample's value. Thus two accepted successive samples in the chain are not independent of each other.

f)

When a proposal is rejected the independence sampler, samples the previous sample again:

$$[x_1, x_1, x_3, x_4, x_4]$$

g)

The rejection sampler compares the volumes between the proposal approximate distribution and the unnormalized true distribution. With higher dimensionality the acceptance ratio will become exponentially small. **TODO FINISH WRITING**

The importance sample importance sampler: no
independence sampler: yes

Problem 2.

We have $x \sim \mathcal{N}(x|\mu, \tau^{-1})$, $\mu \sim \mathcal{N}(\mu|\mu_0, s_0)$ and $\tau \sim \text{Gamma}(\tau|a, b)$.

To gibbs sample from the posterior $p(\mu, \tau|x)$ we iteratively sample from the conditionals $p(\mu|\tau, x)$ and $p(\tau|\mu, x)$.

$$
\begin{aligned}
p(\mu|\tau, x) &= \mathcal{N}\left(\mu| \frac{\tau^{-1}}{s_0 + \tau^{-1}} \cdot \mu_0 + \frac{s_0}{s_0 + \tau^{-1}} \cdot x, \frac{1}{\frac{1}{s_0} + \frac{1}{\tau^{-1}}}\right) \\
p(\tau|\mu, x) &= \text{Gamma}(\tau|a + \frac{1}{2}, b + \frac{1}{2}(x - \mu)2)
\end{aligned}
$$

(using Bishop 2.141/142 and 2.150/151)

Problem 3.

1.

2.

3.

Problem 4.

We have $p(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{i=1}^{D} \mu_i^{x_i}(1 - \mu_i)^{1-x_i}$, $\mu_i \in [0, 1]$, $x_i \in \{0, 1\}$.

a)

$$\mathbb{E}[\boldsymbol{x}] = \boldsymbol{\mu}$$

b)

$$\text{cov}(\boldsymbol{x}) = \text{diag}(\boldsymbol{\mu}^T(1 - \boldsymbol{\mu}))$$

c)

Now we got an mixture: $p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}|\boldsymbol{\mu}_k)$. And therefore now:

$$\mathbb{E}[\boldsymbol{x}] = \sum_{k=1}^{K} \pi_k \boldsymbol{\mu}_k$$

d)

$$
\begin{aligned}
\log p(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\pi}) &= \log \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{\mu},\boldsymbol{\pi}) \\
&= \log \prod_{n=1}^{N} \sum_{k=1}^{K} \left(\pi_k p(\boldsymbol{x}_n|\boldsymbol{\mu}_k)\right) \\
&= \log \prod_{n=1}^{N} \sum_{k=1}^{K} \left(\pi_k \prod_{i=1}^{D} \mu_{ki}^{x_{ni}} (1-\mu_{ki})^{1-x_{ni}}\right) \\
&= \sum_{n=1}^{N} \log \left[\sum_{k=1}^{K} \left(\pi_k \prod_{i=1}^{D} \mu_{ki}^{x_{ni}} (1-\mu_{ki})^{1-x_{ni}}\right)\right]
\end{aligned}
$$

e)

A standard maximum likelihood approach would not work here as we have need to compute the logarithm of the sum. We do not have a closed-form solution and the computation of this logarithm is difficult.

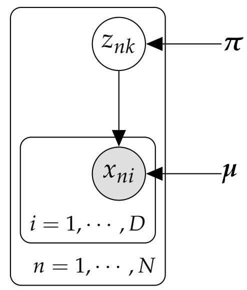
f)

For our variationl approach we have: $p(\boldsymbol{x}_n, \boldsymbol{z}_n|\boldsymbol{\mu},\boldsymbol{\pi}) = p(\boldsymbol{z}_n|\boldsymbol{\pi})p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\mu}) = \prod_{k=1}^{K} \pi_k^{z_{nk}} p(\boldsymbol{x}_n|\boldsymbol{\mu}_k)^{z_{nk}}$.

The data log likelihood becomes:

$$
\begin{aligned}
\log p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\mu},\boldsymbol{\mu}) &= \log \prod_{n=1}^{N} p(\boldsymbol{z}_n|\boldsymbol{\pi})p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\mu}) \\
&= \sum_{n=1}^{N} \log \prod_{k=1}^{K} \left(\pi_k^{z_{nk}} p(\boldsymbol{x}_n|\boldsymbol{\mu}_k)^{z_{nk}}\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \log \left(\pi_k^{z_{nk}} p(\boldsymbol{x}_n|\boldsymbol{\mu}_k)^{z_{nk}}\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \left(z_{nk} \cdot \log \pi_k + z_{nk} \cdot \log p(\boldsymbol{x}_n|\boldsymbol{\mu}_k)\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \left[z_{nk} \cdot \log \pi_k + z_{nk} \cdot \log \left(\prod_{i=1}^{D} \mu_{ki}^{x_{ni}} (1-\mu_{ki})^{1-x_{ni}}\right)\right] \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \left[z_{nk} \cdot \log \pi_k + z_{nk} \cdot \sum_{i=1}^{D} \left(x_{ni} \cdot \log \mu_{ki} + (1-x_{ni}) \cdot \log(1-\mu_{ki})\right)\right]
\end{aligned}
$$

g)



h)

i)

j)