PBM SS2020

1 General Information

- You will write your answers in the dedicated areas in the format you want Latex, pseudo-code, etc.
- Note that if you want to write Latex answers you just have to surround your equation by "(" and ")" (remove the back ticks)
- On top of this you will be able to upload any sheets of computations you used to get to your answer: Just take a picture/scan of it and upload it (there is a specific section in each exercise for it). It is preferable that you separate computations for each section.
- The exam will happen on `a day` from `a time` to `a time`. The time from 16:00 to 16:15 will be considered to be used only to scan and upload eventual drafts that you have.
- Finally and most importantly, you will be asked join a Zoom meeting, turn on your camera and turn off your microphone. Here are the details :
- Join Zoom Meeting `here would be the meeting link`
- You can either ask your questions on the Zoom chat (but do not disclose any answers), if you have a questions that require secrecy, send me a private message.

`things that were not part of the exam are written in monospace`

2 Inference with Gaussian random variables

Suppose we have two random variables U and V which are jointly Gaussian distributed with means $E[U] = a, E[V] = b$ and variances $E\left[U^2\right] = C_u$ and $E\left[V^2\right] = C_v$.

`edit: in the exam it was anounced that these should be VAR[...].`

The covariance is $E[UV] = C_{uv}$ Assume that we observe a noisy estimate

$$Y = V + \varepsilon$$

of V, where ε is a Gaussian noise variable independent of U and of V with $E[\varepsilon] = 0$ and $E\left[\varepsilon^2\right] = \sigma^2$.

The following formulas could be helpful: The inverse of the matrix x

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

is given by

$$\mathbf{S}^{-1} = \frac{1}{\det \mathbf{S}} \begin{pmatrix} S_{22} & -S_{12} \\ -S_{21} & S_{11} \end{pmatrix}$$

The determinant iss

$$\det \mathbf{S} = S_{11}S_{22} - S_{12}S_{21}$$

The one dimensional Gaussian density for a random variable with mean $E[x] = \mu$ and variance $\sigma^2 = E[x-\mu]^2$ is given by

$$p\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The multivariate Gaussian density for a random vector $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$ and covariance matrix $\mathbf{S}$ is given by

$$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathbf{S}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{S}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

Note, that $S_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$

2.1 10 Points

Calculate the mean vector m and the covariance matrix S for the random vector (U, Y).

2.2 5 Points

Compute the joint density $p(U, Y \mid m, S)$ for the values
$C_u = 3, C_v = 1, a = 0, b = 1$ $C_{uv} = \sqrt{2}$ and $\sigma^2 = 1$

2.3 10 Points

Compute $E[U \mid Y]$ and $\text{VAR}[U \mid Y]$ for $Y = 2$?

3 EM algorithm for a geometric mixture model

Consider a mixture model for a integer valued random variable $n \in \{0, 1, 2, \ldots\}$ given by the distribution

$$P(n \mid \mathbf{q}) = \sum_{j=1}^{M} P(j) P\left(n \mid q_j\right)$$

where the component probabilities $P\left(n \mid q_j\right)$ are geometric distributions

$$P(n \mid q) = q(1-q)^n$$

Based on a data set of N i.i.d. $\sim$ samples $D = (n_1, n_2, \ldots, n_N)$ we want to estimate the parameters $\mathbf{q} = (q_1, \ldots, q_M, P(1), \ldots, P(M))$ of this mixture model

3.1 8 Points

Derive an expression for the Maximum Likelihood estimate of q_1 for $M = 1$, where all observations come from the same geometric distribution.

3.2 6 Points

For $M > 1$ the maximum likelihood estimates of the parameters are to be determined using an EM algorithm.

For the E-step, compute

$$\mathcal{L}\left(\boldsymbol{q}, \boldsymbol{q}_t\right) = -\sum_{i=1}^{N} \sum_{j=1}^{M} P_t\left(j \mid n_i\right) \ln\left(P\left(n_i \mid q_j\right) P(j)\right)$$

where $P_t\left(j \mid n_i\right)$ is the responsibility of component j for generating data point n_i, computed with the current values of the parameters.

For the M-step, minimise $\mathcal{L}$ with respect to q_j and give an explicit expression for the EM-update of q_j

You don't have to compute the update of $P(j)$

4 Bayes inference and Gibbs sampler

Consider the geometric distribution

$$P(n \mid q) = q(1-q)^n$$

for $n = 0, 1, 2, \ldots$

4.1 5 Points

Find the conjugate prior density of the geometric distribution.

4.2 5 Points

Assume a data set of i.i.d. $\sim$ samples $D = (n_1, n_2, \ldots, n_N)$ drawn from the geometric distribution and a beta distributed prior $p(q) = \text{Beta}(a, b)$.

Compute the posterior density $p(q \mid D)$

4.3 5 Points

What is the MAP value of q?

5 Outlier detection with Gibbs sampling

We assume a data set $D = (n_1, \ldots, n_N)$, where observations are drawn with known probability $1 - c = 0.9$ from a geometric distribution ('regular' observations) with unknown parameter q, i.e.

$$P_0(n \mid q) = q(1-q)^n$$

But with probability $c = 0.1$ data points are 'outliers'. In this case, the distribution of n is assumed to be Poisson with a known parameter l

$$P_1(n) = \exp(-l)\frac{l^n}{n!}$$

We assume a conjugate prior for the parameter q :

$$q \sim \text{Beta}(a, b)$$

5.1 10 Points

Introduce for each data point a latent indicator variable $d_i \in \{0, 1\}$, which decides if a datapoint is regular or an outlier, i.e.

$$d_i = \begin{cases} 1 & \text{if } n_i \text{ is an outlier,} \\ 0 & \text{if } n_i \text{ is a regular data point} \end{cases}$$

Hence, we have

$$P(d_i) = c^{d_i}(1-c)^{1-d_i}$$

Write down the joint distribution of all variables

$$P(D, \boldsymbol{d}, \boldsymbol{q})$$

where $d = (d_1, \ldots, d_N)$ is the vector of indicator variables for each data point.

6 Variational Inference

Assume we have n observations $D = (x_1, \ldots, x_n)$ generated independently from a Gaussian density with unit variance

$$p\left(x_i \mid m, Z\right) = \left(\frac{1}{2\pi}\right)^{1/2} \exp\left[-\frac{1}{2}\left(x_i - Zm\right)^2\right]$$

with a latent variable $Z \in \{0, 1\}$. This means, that we assume that the mean of the Gaussian is either exactly zero ($Z = 0$) or we have some unknown mean which can have any value m We assume that the prior probability of Z is simply given by

$$P(Z = 1) = P(Z = 0) = 1/2$$

and the prior distribution for m is

$$p(m) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{m^2}{2s^2}\right)$$

For all questions you are allowed to use the following results which follow from the derivations given in the lecture:

$$q_1(m) \propto \exp\left[E_Z[\ln p(D, m, Z)]\right]$$
$$q_2(Z) \propto \exp\left[E_m[\ln p(D, m, Z)]\right]$$

6.1 4 Points

Write down the joint probability distribution of all variables

$$p(D, Z, m)$$

6.2 8 Points

We want to find the optimal factorizing approximation $q(m, Z) = q_1(m) q_2(Z)$ which minimises the Kullback-Leibler divergence between q and the posterior $p(m, Z \mid D)$.
Find the optimal distribution $q_1(m)$ and give expressions for its parameters in terms of expectations with respect to $q_2(Z)$ (written as E[Z], $\mathrm{E}[Z^2]$, etc).

6.3 8 Points

Compute the probability $q_2(Z = 1)$ in terms of expectations with respect to $q_1(m)$ (written as $E[m]$,$\mathrm{E}\left[m^2\right]$, etc).