PBM SS2020 - Second Exam Date

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 `a time` to `a time` will be considered to be used only to scan and upload eventual drafts that you have. `(= you have addditional 15 minutes to upload your solutions)`
- 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`

`take the number of points with a grain of salt`

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 $E\left[U^2\right] = C_\alpha$ and $E\left[V^2\right] = C_v$. We also know the expectation $E[UV] = C_{\mathrm{ww}}$ 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

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

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 is

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

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 $x = (x_1, \ldots, x_d)^T$ with mean $\mu = (\mu_1, \ldots, \mu_d)^T$ and covariance matrix S is given by

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

```
the formula above might have an error.
usually the denominator is given with the det S and not |S|.
```

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 $C_u = 3, C_v = 1, a = 0, b = 1$ $C_{\text{uv}} = -\sqrt{2}$ and $\sigma^2 = 1$
(give the explicit expression for the density distribution).

2.3 10 Points

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

3 EM algorithm for a geometric mixture model

Consider a mixture model for a non negative random variable x given by the density

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

where the component probabilities $P\left(x \mid q_f\right)$ are exponential densities

$$P(x \mid q) = q e^{-qx}$$

Based on a data set of N i.i.d. samples $D = (x_1, x_2, \ldots, x_N)$ 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 exponential 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

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

where $P_t\left(j \mid x_i\right)$ is the responsibility of component j for generating data point x_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 exponential density

$$P(x \mid q) = qe^{-qx}$$

for $x \geq 0$

4.1 5 Points

Show that the conjugate prior density for the exponential density is a Gamma density Gamma (a, b) which is given by:

$$p(q \mid a, b) = C(a, b)q^{a-1}e^{-bq}$$

where $C(a, b)$ is a normalising constant.

4.2 5 Points

Assume a data set of i.i.d. samples $D = (x_1, x_2, \ldots, x_N)$ drawn from the exponential density and a Gamma distributed prior $p(q) = \text{Gamma}(a, b)$.

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 $D = (x_1, \ldots, x_N)$, where observations are drawn with known probability $1 - c = 0.9$ from an exponential density (**regular** observations) with unknown parameter q, i.e.

$$P_0(x \mid q) = qe^{-qx}$$

but with probability $c = 0.1$ data points are outliers. In this case, the distribution of x is assumed to be of the form

$$P_1(x) = Cx^2e^{-gx}$$

where C is a constant and with a known parameter g.
We assume a conjugate prior for the parameter q :

$$q \sim \mathrm{Gamma}(a, b)$$

5.1 10 Points

We 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.

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

Hence, we have

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

Write down the joint distribution of all variables

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

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

5.2 ? Points

To perform Bayesian inference we want to use a Gibbs sampler. Compute the necessary conditional densities

$$p(d_i \mid q, D, \mathbf{d} \setminus i) \text{ for } i = 1, \ldots, N$$
$$p(q \mid D, \mathbf{d})$$

`This was missing in the first exam date.`

6 Variational Inference

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

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

with a latent variable $Z \in \{-1, 1\}$. This means, that we assume that the unknown mean of the Gaussian can only be 1 and we have an unknown precision (inverse variance) given by λ. We assume that the prior probability of Z is simply given by

$$P(Z = \pm 1) = \frac{1}{2}$$

and the prior distribution for λ is

$$p(\lambda) = \gamma e^{-\gamma\lambda}$$

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

$$q_1(\lambda) \propto \exp\left[E_Z[\ln p(D, \lambda, Z)]\right]$$
$$q_2(Z) \propto \exp\left[E_\lambda \mid \ln p(D, \lambda, Z)] \mid\right]$$

6.1 4 Points

Write down the joint probability distribution of all variables

$$p(D, \lambda, Z)$$

6.2 8 Points

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

6.3 8 Points

`Unknown. Probably very similar to the other SS2020 exam or maybe there was no 6.3 but instead a 5.2 (which was missing in the first exam date).`

7 Grades

This table has been roughly the same for the last few years so it should be the same for the next few years.

>=	Grade
0	5
35	4
40	3,7
45	3,3
50	3
55	2,7
60	2,3
65	2
70	1,7
75	1,3
80	1