HY673 - Spring 2024 Problem Set #1
Yannis Pantazis (12/03/2024)

Instructions

- **Due date:** Tuesday March 26th, 2023
- Submission via e-mail to the class account: hy673@csd.uoc.gr
- Provide one file with the written solutions.
- Provide one folder with code.
  - The name of each file in the folder should indicate the respective exercise.
  - Your code should run on colab.
- All assignments in this course are individual, not group, assignments. You may freely discuss homework assignments with your fellow classmates. The final solutions, however, must be written entirely on your own. This includes programming assignments.
- You are allowed to use Generative Al Tools such as ChatGPT for help with homework assignments for grammatical corrections only. To maintain academic integrity, students must disclose any use of Al-generated material.

Problem 1 (Change of variable formula)**.** *Let $X \sim Exp(\lambda)$ be an exponential random variable with parameter λ and let $Y = X^2$.*

(a) Calculate analytically the probability density function (pdf), $f_Y(y)$, of Y using the change of variable formula.

(b) Compute the histogram of the dataset $\{y_i = g(x_i) : x_i \sim Exp(0.5)\}_{i=1}^n$ with $n = 100, 1000$ & 10^4. Plot in the same figure and compare the estimated histogram with $f_Y(y)$ from (a). What do you observe as n increases?

(c) Repeat (b) using the dataset $\{y_i = F_Y^{-1}(u_i) : u_i \sim \mathcal{U}(0,1)\}_{i=1}^n$ where $F_Y(y) = \int_{-\infty}^{y} f_Y(z)dz$ is the cumulative distribution function. You are allowed to use the function

```
integrate()
```

of SymPy Python library for the estimation of the indefinite integral.

Problem 2 (Multivariate Gaussian)**.** *Assume that $X = [X_1, X_2, X_3]^T \sim \mathcal{N}(\mu, \Sigma)$ where μ is the mean vector and Σ is the covariance matrix.*

(a) Compute the pdf of $Y = X_2 + X_3$ and the pdf of $Z = [X_1, Y]$ assuming that both pdfs are Gaussians.

(b) Compute the conditional pdf

$$p(x_1|x_2 + x_3 = 0)$$

(c) Numerically validate the result in (b) for $\mu = [0, 1, -1.5]^T$ *and*

$$\Sigma = \begin{bmatrix} 1 & 0.5 & -0.8 \\ 0.5 & 0.9 & -0.7 \\ -0.8 & -0.7 & 1.1 \end{bmatrix} .$$

To do so, generate $n = 10^5$ *samples from the 3-dimensional pdf, define a stride via* $-\epsilon < x_2 + x_3 < \epsilon$ *for a small value of* ϵ *and then create a histogram for* x_1 *using only the samples that fall inside the stride. Compare the computed histogram with the solution in (b). What percentage of the samples fall in the stride for* $\epsilon = 0.1$ *and how many when it is set to* $\epsilon = 0.01$*?*

Problem 3 (Maximum likelihood estimation)**.** *Generate and infer the parameters of an autoregressive (AR) process.*

(a) Simulate an AR(1) *process which is given by the formula*

$$x_t = a_0 + a_1 x_{t-1} + w_t \, , \quad t = 0, 1, 2, ..., T-1$$

where w_t *is white noise (i.e.,* $w_t \sim \mathcal{N}(0, \sigma^2)$ *for all* t *and* w_t *is independent of* $w_{t'}$ *for all* t, t' *with* $t \neq t'$*),* $\sigma = 1.0$*,* $a_0 = 2.0$*,* $a_1 = -1.5$*,* $x_{-1} = 0$ *and* $T = 1000$*.*

(b) Write down the log-likelihood of the above AR(1) *process for the parameter vector* $\theta = [a_0, a_1]^T$*.*

(c) Compute analytically and then numerically using the simulated process from (a), the maximum likelihood estimator. Plot the mean squared error between the numerically estimated $\hat{\theta}_{MLE}$ *and the ground truth as a function of* T*.*

Problem 4 (Gaussian Mixture Model (GMM) with prior)**.** *(a) You will derive the Expectation-Maximization (EM) algorithm when prior knowledge regarding the mean values is available. Let* $\left\{\pi, \{\mu_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K\right\}$ *be the parameters of a GMM model with* K *Gaussians and data dimension* d*. Moreover, assume that each* μ_k *is independently sampled from a Gaussian prior,* $\mu_k \sim \mathcal{N}(\mu_{0k}, \lambda^{-1} I_d)$*,* $k = 1, ..., K$ *where* μ_{0k} *is the prior mean vector while* λ *is the inverse variance and it is interpreted as the strength of the prior (e.g., larger values for* λ *implies stronger prior). We assume no prior information regarding the weights,* π *and the covariance matrices,* $\{\Sigma_k\}_{k=1}^K$*.*
Repeat the derivation steps of the EM algorithm starting from the maximization of the logarithm of the a posteriori distribution

$$p\big(\pi, \{\mu_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K | x\big) \propto p\big(x | \pi, \{\mu_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K\big) \times p\big(\{\mu_k\}_{k=1}^K\big)$$

where $p(\{\mu_k\}_{k=1}^K)$ is the (Gaussian) prior distribution for the mean vectors.

<u>Hint:</u> Only the formula for the mean vectors will be different.

(b) Generate $n = 1000$ samples from a GMM with $K = 3$ components using the ancestral sampling algorithm (see Lecture 4). The mean vectors of the three equiprobable Gaussian components are $\mu_1 = [0, 1]^T$, $\mu_2 = [1, -0.5]^T$ and $\mu_3 = [0, 0]^T$ while the respective covariance matrices being

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.9 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1.1 \end{bmatrix}, \quad \textit{and} \qquad \Sigma_3 = \begin{bmatrix} 1.5 & 1.3 \\ 1.3 & 1 \end{bmatrix}$$

(c) Use the equations derived in (a) and the data from (b) to estimate the parameters of the GMM. Consider three cases:
i) Few data with strong correct prior (e.g., $n \approx 100$ or less, $\mu_{0k} \approx \mu_k$ and $\lambda = O(10^3)$),
ii) Few data with strong wrong prior (e.g., $n \approx 100$ or less, $\mu_{0k} \approx \mu_k + 1$ and $\lambda = O(10^3)$),
iii) Many data with strong wrong prior (e.g., $n \approx 10^4$, $\mu_{0k} \approx \mu_k + 1$ and $\lambda = O(10^3)$).

Problem 5 (Evidence lower bound (ELBO))**.** *(a) Let $p(x, z)$ be the joint pdf, $p(x)$ be the marginal pdf (or evidence) and $p(z|x)$ be the posterior pdf. Assume also another conditional pdf denoted by $q(z|x)$. For all x, prove that*

$$\log p(x) = \mathbb{E}_{q(z|x)}\Big[\log \frac{p(x,z)}{q(z|x)}\Big] + D_{KL}\big(q(z|x)\,||\,p(z|x)\big)$$

where $D_{KL}(\cdot\,||\,\cdot)$ denotes the Kullback-Leibler divergence.

(b) Using the above formula from, prove the evidence lower bound for the GMM case which reads (see also slides 18 & 19 in Lecture 4):

$$\log p_\theta(x) \geq \mathbb{E}_{p_{\theta^{old}}(z|x)}\Big[\log p_\theta(x,z)\Big] - \mathbb{E}_{p_{\theta^{old}}(z|x)}\Big[\log p_{\theta^{old}}(z|x)\Big]$$