HY673 - Assignment #1

Yannis Pantazis

Monday February 20th, 2023

Instructions

- **Due date:** Monday March 6th, 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 (e.g., ex_2c.py or exercise_2c.py).

Exercise 1: Change of variables

(a) Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = e^X =: g(X)$. Compute analytically the probability density function (pdf) of Y using the change of variable formula.

(b) Compute the histogram of the dataset $\{y_i = g(x_i) : x_i \sim \mathcal{N}(1,4)\}_{i=1}^n$ with $n = 100, 1000$ & 10^4. Compare the estimated histogram with $P_Y(y)$ from (a). Write down what you observe as n increases.

(c) Let $U_1, U_2 \sim \mathcal{U}(0,1)$ two independent random variables and define $Y = -\lambda_1^{-1}\log(U_1) - \lambda_2^{-1}\log(U_2)$ with $\lambda_1, \lambda_2 > 0$. Compute analytically the pdf of Y.

(d) Repeat (b) but for (c) with $\lambda_1 = 1$ and $\lambda_2 = 2$.

Exercise 2: Multivariate Gaussian

(a) Let X_1, X_2 be two *dependent* Gaussians with $\mathbb{E}[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$ for $i = 1, 2$ and $\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \sigma_{12}$. First show that the sum $Y = X_1 + X_2$ is Gaussian and then compute the pdf of Y.

(b) Compute $p(x_1 + x_2|x_3)$ given that $p(x_1, x_2, x_3) \equiv \mathcal{N}(\mu, \Sigma)$, $\mu \in \mathbb{R}^3$ and $\Sigma \in \mathbb{R}^{3\times 3}$.

(c) Using $p(x_1, x_2|x_3)$ as an intermediate step, numerically validate through histogram comparisons the result in (b) for $\mu = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & -0.5 & 0.8 \\ -0.5 & 1 & -0.8 \\ 0.8 & -0.8 & 1 \end{bmatrix}$.

Exercise 3: MLE - Linear case

(a) Let $x_1, ..., x_n \in \mathbb{R}$ be a given dataset. Assume the parametric model that generates the samples to be $p_\theta(x) \equiv \mathcal{N}(0, \theta)$ (i.e., θ corresponds to the variance σ^2). Compute analytically the MLE solution, $\hat{\theta}_{MLE}$.

(b) Generate samples $x_i \sim \mathcal{N}(0, \theta^*)$ with $\theta^* = 2$ and create one realization of the dataset. For $n = 10, 100$ & 10^3, plot the log-likelihood as a function of θ. What do you observe (write down at least two observations)?

(c) Repeat (b) averaging over 100 realizations with new data at each realization and plot the average log-likelihood function (again, as a function of θ). For each realization, compute also $\hat{\theta}_{MLE}(n)$ and plot the respective histogram. Comment on the shape of the average log-likelihood function and the shape of the histogram of $\hat{\theta}_{MLE}(n)$.

(d) Compute analytically the Fisher information

$$\mathcal{I}(\theta) := \mathbb{E}_{p_\theta}\left[\left(\frac{d}{d\theta}\log p_\theta(x)\right)^2\right] = -\mathbb{E}_{p_\theta}\left[\frac{d^2}{d\theta^2}\log p_\theta(x)\right].$$

Exercise 4: MLE - Nonlinear case

(a) Let $p_\theta(x) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2}$ be the 1D logistic distribution with $\theta = \begin{bmatrix}\mu \\ s\end{bmatrix}$ where μ corresponds to the mean parameter while s corresponds to the scaling parameter. Compute the log-likelihood function and its partial derivatives with respect to μ and s.

(b) For $x_1, ..., x_n$ generated from the logistic distribution with $\mu^* = 1$ and $s^* = 0.5$, plot the 2D contour of the log-likelihood function for $n = 100, 1000$ and 10^4.

(c) For the data generated in (b), estimate iteratively the MLE solution, $\hat{\theta}_{MLE}$, using the gradient ascent algorithm. How mean squared error behaves as a function of n?

Exercise 5: Mixture Model

(a) Let $X = [X_1, ..., X_d]^T$ with each $X_j \sim \text{Bernoulli}_1(\mu_j)$ independent to each other be a multivariate Bernoulli distribution. Recall that $\mu_j \in [0, 1]$. The probability mass function (pmf) of X is given by

$$\text{Bernoulli}_d(x|\mu) = \prod_{j=1}^{d} \mu_j^{x_j}(1-\mu_j)^{1-x_j}.$$

A Bernoulli mixture model (BMM) with K components and mixing coefficient vector $\pi = [\pi_1, ..., \pi_K]^T$ is given by

$$p_\theta(x) = \sum_{k=1}^{K} \pi_k \text{Bernoulli}_d(x|\mu_k),$$

with $\mu_k = [\mu_{k1}, ..., \mu_{kj}, ..., \mu_{kd}]^T \in \mathbb{R}^d$ and $\theta = \{\pi_k, \mu_k\}_{k=1}^K$.

(a) Compute the log-likelihood of the BMM as well as the likelihood assuming that the latent variable (indicating from which component a sample is drawn) is given.

(b) Derive the expectation step along with the formula for the responsibilities and then the maximization step of the EM algorithm.

(c) Implement the EM algorithm for the BMM and train it on one of the MNIST digit images. MNIST images have to be down-sampled to 14x14 dimension and binarize. You may adapt available code that implements the EM algorithm for the GMM. Generate new synthetic images using the trained BMM.