6^{th} Homework

Chalkiopoulos Georgios | $p3352124$

February 25, 2022

Exercise 1.

$$p(x) = \theta^2 x \exp(-\theta x) u(x),\ u(x) = 1(0)\ if\ x \geq 0 (< 0) \tag{1.1}$$

(a) Prove the the ML estimate of θ is $\theta_{ML} = \dfrac{2N}{\sum_{i=1}^{N} x_i}$.

The Log-likelihood function is:

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \ln\ p(X;\boldsymbol{\theta}) = lnp(x_1, \ldots, x_N) \\
&= \sum_{i=1}^{N} \ln\ p(x_i;\boldsymbol{\theta}) \\
&= \sum_{i=1}^{N} \ln\ (\theta^2 \exp(-\theta x_i) u(x_i)) \\
&= \sum_{i=1}^{N} \ln\ (\theta^2 \exp(-\theta x_i) u(x_i)) \\
&= \sum_{i=1}^{N} (2 \ln\ \theta + \ln\ \exp(-\theta x_i) + ln\ u(x_i)) \\
&= \sum_{i=1}^{N} (2ln\ \theta - \theta x_i + \ln\ u(x_i))
\end{aligned} \tag{1.2}$$

Taking the partial derivative, in regards to θ it is:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{ML} = arg\, max_{\theta} p(Y;\theta) \Rightarrow \frac{\partial L(\boldsymbol{\theta})}{\partial \theta} = 0 \Rightarrow \\
\frac{\partial \sum_{i=1}^{N} (2ln\ \theta - \theta x_i + \ln\ u(x_i))}{\partial \theta} = 0 \Rightarrow \\
\sum_{i=1}^{N} (\frac{\partial\ 2ln\theta}{\partial \theta} - \frac{\partial \theta x_i}{\partial \theta}) = 0 \Rightarrow \\
\sum_{i=1}^{N} \frac{2}{\theta} = \sum_{i=1}^{N} x_i \\
\frac{1}{\hat{\boldsymbol{\theta}}_{ML}} \sum_{i=1}^{N} 2 = \sum_{i=1}^{N} x_i \\
\hat{\boldsymbol{\theta}}_{ML} = \frac{2N}{\sum_{i=1}^{N} x_i}
\end{aligned} \tag{1.3}$$

(b) For N=5 and X = {2, 2.2, 2.7, 2.4, 2.6} estimate $\hat{\boldsymbol{\theta}}_{ML}$ and determine $\hat{p}(x)$ for x = 2.3 and 2.9.

Using (1.3) it is:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{ML} &= \frac{2N}{\sum_{i=1}^{N} x_i} \\
&= \frac{10}{2 + 2.2 + 2.7 + 2.4 + 2.6} \\
&= 0.84
\end{aligned}$$

Therefore:

$$\hat{p}(x) = \begin{cases} 0.235, & x = 2.3 \\ 0.179, & x = 2.9 \end{cases}$$

Exercise 2.

$$p(x;\theta) = 2\theta x \exp(-\theta x^2)u(x),\ u(x) = 1(0)\ if\ x \geq 0(< 0) \tag{2.1}$$

(a) Compute the MAP estimate of the parameter θ

It is:

$$\begin{aligned}
\ln\ p(\theta)\ p(x|\theta) &= \ln\left(\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right)\prod_{i=1}^{N} 2\theta x\exp(-\theta x^2)u(x)\right) \\
&= \ln\frac{1}{\sqrt{2\pi}\sigma_0} - \frac{(\theta-\theta_0)^2}{2\sigma_0^2} + N\ln 2\theta + \ln\sum_{i=1}^{N} x_i - \theta\sum_{i=1}^{N} x_i^2 \quad \Rightarrow \\
\frac{\partial \ln\ p(\theta)\ p(x|\theta)}{\partial\theta} &= -\frac{(\theta-\theta_0)}{\sigma_0^2} + \frac{N}{\theta} - \sum_{i=1}^{N} x_i^2 = 0 \quad \Rightarrow \\
-\frac{\theta}{\sigma_0^2} + \frac{\theta_0}{\sigma_0^2} + \frac{N}{\theta} &= \sum_{i=1}^{N} x_i^2 \quad \Rightarrow \\
-\frac{\theta}{\sigma_0^2} + \frac{N}{\theta} &= \sum_{i=1}^{N} x_i^2 - \frac{\theta_0}{\sigma_0^2} \quad \Rightarrow
\end{aligned}$$

We, therefore, have to solve the following polynomial:

$$\theta^2 + \left(\sigma_0^2\sum_{i=1}^{N} x_i^2 - \theta_0\right)\theta - \sigma_0^2 N = 0$$

The quadratic formula is written as:

$$\theta_{1,2} = \frac{\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i^2 \pm \sqrt{(\sigma_0^2\sum_{i=1}^{N} x_i^2 - \theta_0)^2 + 4N\sigma_0^2}}{2} \tag{2.2}$$

Given that we want to estimate the parameter θ of the Rayleigh distribution, we should only keep $\theta_{1,2}$ in which $\theta > 0$. It is easy to prove that this applies for the following root:

$$\theta_{MAP} = \frac{\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i^2 + \sqrt{(\sigma_0^2\sum_{i=1}^{N} x_i^2 - \theta_0)^2 + 4N\sigma_0^2}}{2}$$

b) How this estimate becomes for the case were:

(i) $N \to \infty$ It is:

$$\theta_{MAP} = \frac{-\frac{\sigma_0^2 \sum_{i=1}^{N} x_i^2}{N} + \frac{\theta_0}{N} + \sqrt{(\frac{\sigma_0^2 \sum_{i=1}^{N} x_i^2}{N} - \frac{\theta_0}{N})^2 + 4\sigma_0^2}}{\frac{2}{N}}$$

In this case $\theta_{MAP} \to \infty$.

(ii) $\sigma_0^2 >>$ We write θ_{MAP} as:

$$\theta_{MAP} = \frac{-\sum_{i=1}^{N} x_i^2 + \frac{\theta_0}{\sigma_0^2} + \sqrt{(\sum_{i=1}^{N} x_i^2 - \frac{\theta_0}{\sigma_0^2})^2 + 4N}}{\frac{\theta_0}{\sigma_0^2}}$$

In this case $\theta_{MAP} \to \infty$.

(iii) $\sigma_0^2 <<$ We write θ_{MAP} as:

$$\theta_{MAP} = \frac{\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i^2 + \sqrt{(\sigma_0^2 \sum_{i=1}^{N} x_i^2 - \theta_0)^2 + 4N\sigma_0^2}}{2}$$

In this case $\theta_{MAP} \to 0$.

(c) For $N = 5$ and $x_1 = 2$, $x_2 = 2$, $x_3 = 2$, $x_4 = 2$, $x_5 = 2$, $\theta_0 = 1.8$ and $\sigma_0^2 = 1$ estimate the θ_{MAP}. Utilizing this estimate, determine $\hat{p}(x)$ for $x = 2.3$ and $x = 2.9$. Compare the results with those obtained in exercise 4 of Homework 5, where the ML estimate where considered.

1. It is:

$$\begin{aligned}\theta_{MAP} &= \frac{\theta_0 - \sigma_0^2 \sum_{i=1}^{N} x_i^2 + \sqrt{(\sigma_0^2 \sum_{i=1}^{N} x_i^2 - \theta_0)^2 + 4N\sigma_0^2}}{2} \\ &= \frac{-1 \cdot 28.65 + 1.8 + \sqrt{(1 \cdot 28.65 - 1.8)^2 + 4 \cdot 5 \cdot 1}}{2} \Rightarrow \\ &= 0.185\end{aligned}$$

We can now calculate $\hat{p}(x)$:

$$\hat{p}(x) = \begin{cases} 0.319, & x = 2.3 \\ 0.226, & x = 2.9 \end{cases}$$

Exercise 3.

Consider the model $x = \theta = \eta$ and set of measurements $Y = x_1, \ldots, x_N$ which are noisy versions of θ. . Assume that we have prior knowledge about θ saying that it lies close to θ_0. Formulating the ridge regression problem for this case as follows:

$$min_\theta J(\theta) = \sum_{n=1}^{N}(x_n - \theta)^2, \text{subject to } (\theta - \theta_0)^2 \leq \rho$$

Prove that:

$$\theta_{RR} = \frac{\sum_{n=1}^{N} x_n + \lambda\theta_0}{N + \lambda}$$

We will use the Lagrangian Function of the RR for this case, which can be written as:

$$L(\boldsymbol{\theta}) = \sum_{n=1}^{N}(x_n - \theta)^2 + \lambda((\theta - \theta_0)^2 - \rho)$$

We first take the gradient of $L(\theta)$:

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \sum_{n=1}^{N}(x_n - \theta)^2 + \lambda((\theta - \theta_0)^2 - \rho) \Rightarrow \\
\frac{\partial L(\boldsymbol{\theta})}{\partial\theta} &= \frac{\sum_{n=1}^{N}(x_n - \theta)^2 + \lambda((\theta - \theta_0)^2 - \rho)}{\partial\theta} \\
&= -2\sum_{n=1}^{N}(x_n - \theta) + 2\lambda(\theta - \theta_0) \\
&= -2\sum_{n=1}^{N} x_n + 2N\theta + 2\lambda\theta - 2\lambda\theta_0 \Rightarrow \\
\frac{\partial L(\boldsymbol{\theta})}{\partial\theta} &= 0 \Rightarrow \\
0 &= -2\sum_{n=1}^{N} x_n + 2N\theta_{RR} + 2\lambda\theta_{RR} - 2\lambda\theta_0 \Rightarrow \\
\theta_{RR} &= \frac{\sum_{n=1}^{N} x_n + \lambda\theta_0}{N + \lambda}
\end{aligned}$$

Exercise 4.

Consider:

$$p(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.1}$$

Where λ is modeled by a prior Gamma distribution:

$$p(\lambda) = p(\lambda; \alpha, b) = = \begin{cases} \frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda}, & \lambda \geq 0 \\ 0, & \lambda < 0 \end{cases} \tag{4.2}$$

(a) Determine the likelihood $p(Y|\lambda)$, where $Y = \{x_1, \ldots, x_N\}$

The likelihood function is:

$$\begin{aligned} p(Y; \lambda) &= \prod_{i=1}^{N} p(x_1, \ldots, x_N|\lambda) \\ &= \prod_{i=1}^{N} p(x_i; \lambda) \\ &= \prod_{i=1}^{N} \lambda e^{-\lambda x_i} \\ &= \begin{cases} \lambda^N \exp\left(-\sum_{i=1}^{N} \lambda x_i\right), & x \geq 0 \\ 0, & x < 0 \end{cases} \end{aligned} \tag{4.3}$$

(b) Form the product of the prior and the likelihood and determine the MAP estimate of λ.

It is:

$$\hat{\lambda}_{MAP} = argmax_\lambda(p(Y; \lambda)p(\lambda)) \Rightarrow$$

$$p(Y; \lambda)p(\lambda) = \frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda} \lambda^N \exp\left(-\sum_{i=1}^{N} \lambda x_i\right) \Rightarrow$$

$$\ln(\lambda(p(Y;\lambda)p(\lambda)) = \ln\left(\frac{b^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-b\lambda}\lambda^N \exp\left(-\sum_{i=1}^{N}\lambda x_i\right)\right) \Rightarrow$$

$$= \alpha\ln b - \ln\Gamma(\alpha) + (\alpha-1)\ln\lambda - b\lambda + N\ln\lambda - \lambda\sum_{i=1}^{N}x_i \Rightarrow$$

$$\frac{\partial\ln(\lambda(p(Y;\lambda)p(\lambda)}{\partial\lambda} = \frac{(\alpha-1)}{\lambda} - b + \frac{N}{\lambda} - \sum_{i=1}^{N}x_i = 0 \Rightarrow$$

$$\frac{\alpha-1+N}{\lambda} = \sum_{i=1}^{N}x_i + b \Rightarrow$$

$$\hat{\lambda}_{MAP} = \frac{\alpha-1+N}{\sum_{i=1}^{N}x_i + b} \tag{4.4}$$

(c) Give the form of $p(x)$ in terms of the MAP estimate of λ.

Combining (4.1) and (4.4) we get:

$$\hat{p}(x) = \begin{cases} \hat{\lambda}_{MAP}e^{-\hat{\lambda}_{MAP}x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$= \begin{cases} \frac{\alpha-1+N}{\sum_{i=1}^{N}x_i + b}\exp -\frac{\alpha-1+N}{\sum_{i=1}^{N}x_i + b}x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

(d) Determine the posterior distribution for λ: $p(\lambda|Y)$

It is:

$$p(\lambda|Y) = \frac{p(\lambda)p(Y|\lambda)}{p(Y)}$$

$$= \frac{1}{p(Y)}\frac{b^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-b\lambda}\lambda^N \exp\left(-\sum_{i=1}^{N}\lambda x_i\right)$$

$$= \frac{b^\alpha}{p(Y)\Gamma(\alpha)}\lambda^{\alpha-1+N}\exp\left(-(\sum_{i=1}^{N}x_i + b)\lambda\right) \tag{4.5}$$

We observe that (4.5) is a Gamma distribution $C\lambda^r e^{-s\lambda}$ where:

$$C = \frac{b^\alpha}{p(Y)\Gamma(\alpha)}$$
$$r = \alpha - 1 + N \tag{4.6}$$
$$s = \sum_{i=1}^{N} x_i + b \tag{4.7}$$

Thus:

$$p(\lambda|Y) = Gamma(r, s) = \frac{s^r}{\Gamma(r)}\lambda^{r-1}e^{-s\lambda}$$

Comparing the posterior (4.5) with the prior (4.2) we observe that both follow the same distribution, but the posterior has additional information, since the pdf of Y as well as information regarding the data (x_is) are included.

(f) Prove that $p(x|Y)$ is a lomax distribution.

It is:

$$\begin{aligned}
p(x|Y) &= \int p(x|\lambda)p(\lambda|Y)\, d\lambda \\
&= \int_o^\infty \lambda e^{-\lambda x}\, \frac{s^r}{\Gamma(r)}\lambda^{r-1}e^{-s\lambda}\, d\lambda \\
&= \frac{s^r}{\Gamma(r)} \int_o^\infty \lambda^r e^{-(x+s)\lambda}\, d\lambda \\
&= \frac{s^r}{\Gamma(r)} \frac{\Gamma(r+1)}{(x+s)^{r+1}} \\
&= \frac{s^r}{\Gamma(r)} \frac{r\Gamma(r)}{(x+s)^{r+1}} \\
&= \frac{rs^r}{(x+s)^{r+1}} \\
&= Lomax(r, s)
\end{aligned} \tag{4.8}$$

Combining (4.6), (4.7) and (4.8) we finally have that:

$$p(x|Y) = Lomax\left(\alpha - 1 + N, \sum_{i=1}^{N} x_i + b\right)$$

For the following, assume that Y={2.8,2.4,2.9,2.6,2.1,2.2}, $\alpha = 2$ and $b = 2$.

We calculate the sum of x_i as:

$$\sum_{i=1}^{6} x_i = 15$$

(g) Write down the $p(x)$ of (c) for the above Y.

Using $p(x)$ calculated in (c) it is:

$$\begin{aligned} p(x) &= \begin{cases} \frac{\alpha - 1 + N}{\sum_{i=1}^{N} x_i + b} \exp - \frac{\alpha - 1 + N}{\sum_{i=1}^{N} x_i + b} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ &= \begin{cases} \frac{2 - 1 + 6}{15 + 2} \exp - \frac{2 - 1 + 6}{15 + 2} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ &= \begin{cases} \frac{7}{17} \exp - \frac{7}{17} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \end{aligned} \tag{4.9}$$

(h) Write down the $p(x|Y)$ of (f) for the above Y.

Using the $p(x|Y)$ calculated in (f) it is:

$$p(x|Y) = Lomax(7, 17) = \frac{2.87e9}{(x + 17)^7}(0), X \geq 0(x < 0) \tag{4.10}$$

(i) Compute $p(x)$ and $p(x|Y)$ for $x = 2.5$

Using (4.9) and (4.10) we get:

$$\begin{aligned} p(2.5) &= 0.147 \\ p(2.5|Y) &= 0.137 \end{aligned}$$

Exercise 5.

Consider:

$$p(x) = \sum_{j=1}^{m} P_j p(x|j), \quad \sum_{j=1}^{m} P_j = 1, \quad \int_{-\infty}^{+\infty} p(x|j) = 1$$

Where $m = 3$ and P_j, j = 1,2,3 are the apriori probabilities of the pdfs $p(x|j)$, involved in the definition of p(x). In the "parameter updating" part of the EM-algorithm we need to solve the problem:

$$[P_1, P_2, P_3] = argmax_{[P_1,P_2,P_3]} \sum_{i=1}^{N} \sum_{j=1}^{3} P(j|x_i) \ \ln P_j$$

subject to $\sum_{j=1}^{3} P_j = 1$ for fixed $P(j|x_i)$. Prove that the solution to the above problem is:

$$P_j = \frac{1}{N} \sum_{i=1}^{N} P(j|x_i)$$

1. The Lagrangian function is:

$$L(P_1, P_2, P_3) = \sum_{i=1}^{N} \sum_{j=1}^{3} P(j|x_i) \ \ln P_j + \lambda(\sum_{j=1}^{m} P_j - 1) \tag{5.1}$$

2. Focusing on P_1 we will solve the equation: $\frac{\partial L(P_1, P_2, P_3)}{\partial P_1}$:

$$\begin{aligned}
&\frac{\partial L(P_1, P_2, P_3)}{\partial P_1} = 0 \Rightarrow \\
&\frac{\partial \sum_{i=1}^{N} \sum_{j=1}^{3} P(j|x_i) \ \ln P_1}{\partial P_1} + \frac{\partial \lambda(\sum_{j=1}^{m} P_j - 1)}{\partial P_1} = 0 \Rightarrow \\
&\frac{\sum_{i=1}^{N} P(1|x_i)}{P_1} + \lambda = 0 \Rightarrow \\
&P_1 = -\frac{\sum_{i=1}^{N} P(1|x_i)}{\lambda}
\end{aligned} \tag{5.2}$$

P_2 and P_3 are calculated in the same manner.

3. Using $\sum_{j=1}^{m} P_j = 1$ it is:

$$\sum_{j=1}^{m} P_j = 1 \Rightarrow -\frac{\sum_{i=1}^{N} P(1|x_i)}{\lambda} - \frac{\sum_{i=1}^{N} P(2|x_i)}{\lambda} - \frac{\sum_{i=1}^{N} P(3|x_i)}{\lambda} = 1 \Rightarrow$$

$$\lambda = -\sum_{j=1}^{m}\sum_{i=1}^{N} P(j|x_i) \tag{5.3}$$

4. Substituting (5.3) to (5.2) we have:

$$\begin{aligned} P_j &= -\frac{\sum_{i=1}^{N} P(j|x_i)}{\lambda} \\ &= \frac{\sum_{i=1}^{N} P(j|x_i)}{\sum_{j=1}^{m}\sum_{i=1}^{N} P(j|x_i)} \\ &= \frac{\sum_{i=1}^{N} P(j|x_i)}{\sum_{i=1}^{N}\sum_{j=1}^{m} P(j|x_i)} \end{aligned} \tag{5.4}$$

Finally, using that $\sum_{j=1}^{m} P(j|x_i) = 1$ (5.4) is now written as:

$$\begin{aligned} P_j &= \frac{\sum_{i=1}^{N} P(j|x_i)}{\sum_{i=1}^{N}\sum_{j=1}^{m} P(j|x_i)} \\ &= \frac{\sum_{i=1}^{N} P(j|x_i)}{\sum_{i=1}^{N} 1} \Rightarrow \end{aligned}$$

$$P_j = \frac{1}{N}\sum_{i=1}^{N} P(j|x_i)$$

Exercise 6.

Consider again the setup of exercise 5, where now $p(x|j)$'s are normal distributions with means $\boldsymbol{\mu}_j$ and fixed covariance matrices $\Sigma_j, j = 1, \ldots, \mu$. Prove that the solution of the optimization problems

$$\boldsymbol{\mu}_j = \text{argmax}_{\boldsymbol{\mu}_j} \sum_{i=1}^{N} P(j|x_i) \ln \left(p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j) \right), \ j = 1, dots, \mu \tag{6.1}$$

is

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} P(j|\boldsymbol{x}_i)\boldsymbol{x}_i}{\sum_{i=1}^{N} P(j|\boldsymbol{x}_i)} \tag{6.2}$$

Since each pdf is a normal distribution, the term $p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j)$ can be written as:

$$p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j) = \frac{1}{(2\pi)^j |\Sigma_j|^{1/2}} \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j)}{2} \right)$$

Which, as shown in the 3rd assignment, can be written as:

$$p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2} \right)$$

Therefore:

$$\ln p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j) = -\sum_{i=1}^{N} \left(\ln \frac{1}{\sqrt{2\pi\sigma_j^2}} + \frac{(x_i - \mu_j)^2}{2\sigma_j^2} \right)$$

Taking the gradient of (6.1) with respect to $\boldsymbol{\mu}_j$, and setting it to zero it is:

$$\frac{\partial \sum_{i=1}^{N} P(j|x_i) \ln \left(p(\boldsymbol{x}_i|j; \boldsymbol{\mu}_j) \right)}{\partial \boldsymbol{\mu}_j} = 0 \Rightarrow$$

$$\frac{\partial \sum_{i=1}^{N} P(j|x_i) \left(- \left(\ln \frac{1}{\sqrt{2\pi\sigma_j^2}} + \frac{(x_i - \mu_j)^2}{2\sigma_j^2} \right) \right)}{\partial \boldsymbol{\mu}_j} = 0 \Rightarrow$$

$$2\sum_{i=1}^{N} P(j|x_i)\frac{(x_i-\mu_j)}{2\sigma_j^2} = 0 \Rightarrow$$

$$\sum_{i=1}^{N} P(j|x_i)x_i - \sum_{i=1}^{N} P(j|x_i)\mu_j = 0 \Rightarrow$$

$$\sum_{i=1}^{N} P(j|x_i)x_i = \sum_{i=1}^{N} P(j|x_i)\mu_j \Rightarrow$$

$$\mu_j = \frac{\sum_{i=1}^{N} P(j|x_i)x_i}{\sum_{i=1}^{N} P(j|x_i)}$$

Exercise 7.

Consider $X = \{(0,0),(3,0),(12,12),(15,12)\}$ and $p(\boldsymbol{x}) = P_1\ p_1(\boldsymbol{x}) + P_2\ p_2(\boldsymbol{x})$, where:

$$\boldsymbol{\mu}_1(0) = [5,0]^T,\ \boldsymbol{\mu}_2(0) = [6,0]^T,\ P_1(0) = 0.1,\ P_2(0) = 0.9$$

We will run the EM algorithm for the previous problem. Using the initial values,

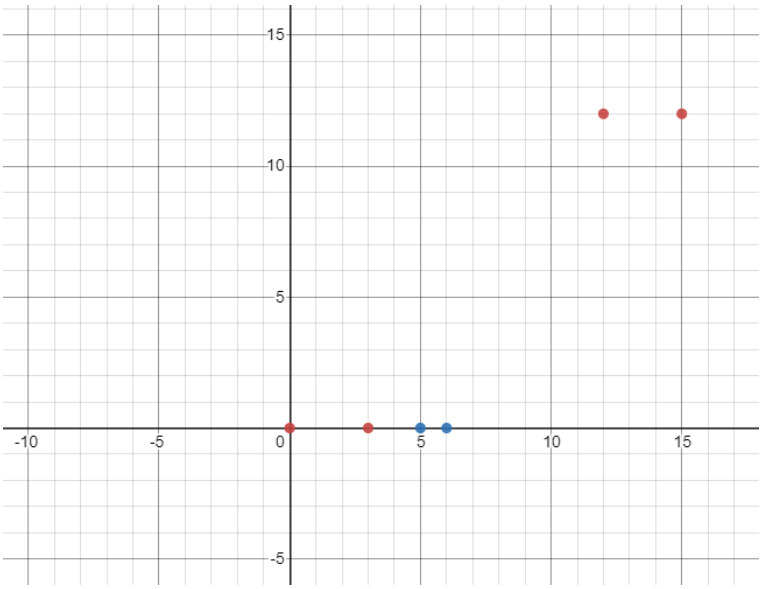


Figure 1: Data set X, with initial with initial mean estimates.

we will calculate the following values:

$$P(1|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|1)P_1}{p(\boldsymbol{x})},\ P(2|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|2)P_1}{p(\boldsymbol{x})}$$

where:

$$p(\boldsymbol{x}) = P_1 \frac{1}{2\pi} \exp(-0.5 \cdot \|\boldsymbol{x} - \boldsymbol{\mu}_1\|^2) + P_2 \frac{1}{2\pi} \exp(-0.5 \cdot \|\boldsymbol{x} - \boldsymbol{\mu}_2\|^2)$$

For each iteration a table will be constructed to show and plot with the results.

Iteration 1.

$$\begin{aligned} \boldsymbol{\mu}_1(0) &= [5,0]^T \quad P_1(0) = 0.1 \\ \boldsymbol{\mu}_2(0) &= [6,0]^T \quad P_2(0) = 0.9 \end{aligned}$$

Regarding x_1 It is:

$$\begin{aligned} p(\boldsymbol{x}) &= 0.1\frac{1}{2\pi}\exp(-0.5\cdot\|\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}5\\0\end{bmatrix}\|^2)+0.9\frac{1}{2\pi}\exp(-0.5\cdot\|\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}6\\0\end{bmatrix}\|^2) \\ &= \frac{0.1}{2\pi}\exp(-0.5\cdot\|\begin{bmatrix}-5\\0\end{bmatrix}\|^2)+0.9\frac{1}{2\pi}\exp(-0.5\cdot\|\begin{bmatrix}-6\\0\end{bmatrix}\|^2) \\ &= 6.15\text{e-}08 \end{aligned}$$

Moreover:

$$P(\boldsymbol{x}|1) = 5.93\text{e-}07,\ P(\boldsymbol{x}|2) = 2.42\text{e-}09$$

Finally:

$$P(1|\boldsymbol{x}) = 0.9645,\ P(2|\boldsymbol{x}) = 0.0355$$

We calculate the remaining x_is and construct the following table:

Table 1: A posteriori probs - 1st Iteration.

	x_1	x_2	x_3	x_4
$P(1\|\boldsymbol{x})$	0.9645	0.5751	0.0002	0.0000
$P(2\|\boldsymbol{x})$	0.0355	0.4249	0.9998	1.0000

Finally we calculate the updated values:

$$\boldsymbol{\mu}_1(1) = \frac{\sum_{i=1}^N P(1|x_i)x_i}{\sum_{i=1}^N P(1|x_i)} = \begin{bmatrix}1.7275\\0.0021\end{bmatrix}/1.5398 = [1.1219, 0.0014]^T$$

And:

$$P_1(1) = \frac{1}{N}\sum_{i=1}^N P(1|x_i) = \frac{1}{4}1.5398 = 0.3850$$

In a similar way we get:

$$\begin{aligned} \boldsymbol{\mu}_1(1) &= [1.1219, 0.0014]^T \quad P_1(1) = 0.3850 \\ \boldsymbol{\mu}_2(1) &= [11.4920, 9.7545]^T \quad P_2(1) = 0.6150 \end{aligned}$$

We plot the points, and see that the means are getting closer to each, of 2, data point groups. The green points correspond to $\boldsymbol{\mu}_1(1)$ and $\boldsymbol{\mu}_2(1)$.

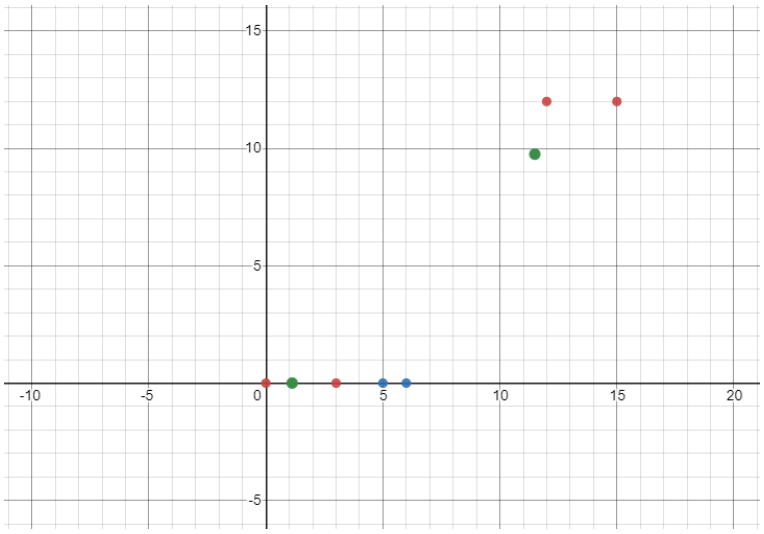


Figure 2: Data set X, with mean estimates after 1^{st} iteration.

Iteration 2.

In a similar way, we create the table for the 2 iteration:

Table 2: A posteriori probs - 2^{nd} Iteration.

	x_1	x_2	x_3	x_4
$P(1\|\boldsymbol{x})$	1.0000	1.0000	0.0000	0.0000
$P(2\|\boldsymbol{x})$	0.0000	0.0000	1.0000	1.0000

The corresponding values are:

$$\boldsymbol{\mu}_1(2) = [1.50, 0.00]^T \quad P_1(2) = 0.5$$
$$\boldsymbol{\mu}_2(2) = [13.5, 12.00]^T \quad P_2(2) = 0.5$$

We plot the points, and see that the means are getting even closer to each, of 2, data point groups. The purple points correspond to $\boldsymbol{\mu}_1(2)$ and $\boldsymbol{\mu}_2(2)$.

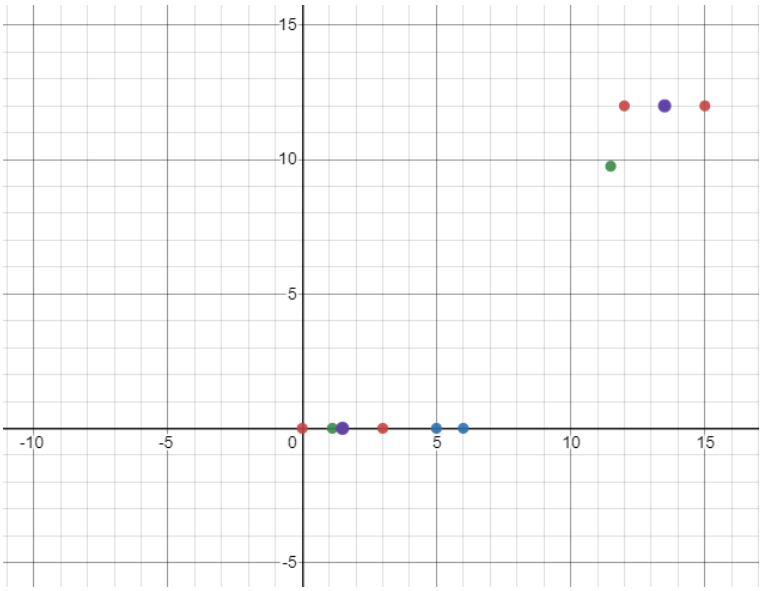


Figure 3: Data set X, with mean estimates after 2$^{\text{nd}}$ iteration.

Iteration 3.

In a similar way, we create the table for the 3 iteration:

Table 3: A posteriori probs - 3$^{\text{rd}}$ Iteration.

	x_1	x_2	x_3	x_4
$P(1\|\boldsymbol{x})$	1.0000	1.0000	0.0000	0.0000
$P(2\|\boldsymbol{x})$	0.0000	0.0000	1.0000	1.0000

The corresponding values are:

$$\boldsymbol{\mu}_1(3) = [1.50, 0.00]^T \quad P_1(3) = 0.5$$
$$\boldsymbol{\mu}_2(3) = [13.5, 12.00]^T \quad P_2(3) = 0.5$$

The estimates did not change since the last iteration, thus we consider a termination criterion to have been met, and we accept those as the final estimates.

Exercise 8

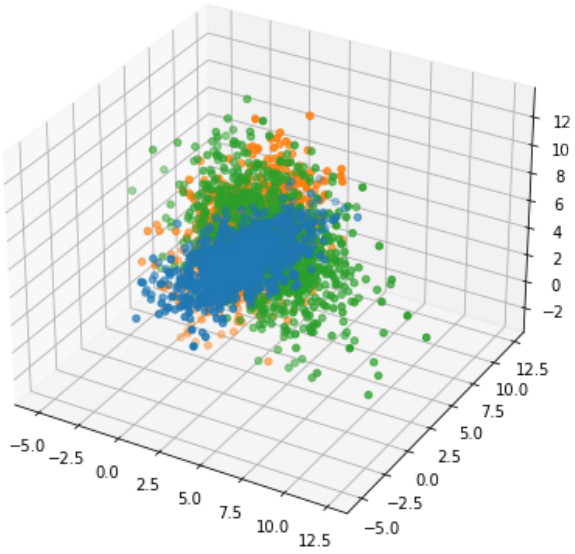
[1]:
```
import scipy.io as sio
import numpy as np
import matplotlib.pyplot as plt
from sklearn.mixture import GaussianMixture
from mpl_toolkits import mplot3d

from sklearn.neighbors import KernelDensity

# for creating a flattened plot
%matplotlib inline
```

[2]:
```
# Load the data
Dataset = sio.loadmat('Dataset.mat')
X1 = Dataset['X1']
X2 = Dataset['X2']
```

8.a Based on $X1$, estimate the values of $p_1(x)$) at the followingpoints:

$x_1 = (2.01, 2.99, 3.98, 5.02)$,$_2 = (20.78, 15.26, 19.38, 25.02)$ $x_3 = (3.08, 3.88, 4.15, 6.02)$

[3]:
```
# Plot the data in pairs to see the distribution
fig = plt.figure(figsize = (10, 7))
ax = plt.axes(projection ="3d")
ax.scatter3D(X1[:,0], X1[:,1], X1[:,2])
ax.scatter3D(X1[:,0], X1[:,2], X1[:,3])
_ = ax.scatter3D(X1[:,1], X1[:,2], X1[:,3])
```



Parametric Approach

We will use the gaussian mixture model approach, using one component (since our data looks like is gathered around one cluster)

[4]:
```
# Parametric approach
gm_1 = GaussianMixture(n_components=1, random_state=0).fit(X1)
mean = gm_1.means_
cov = gm_1.covariances_
mean
```

[4]:
```
array([[1.88364427, 2.94930523, 3.94277602, 4.91643708]])
```

[5]:
```
x_1 = np.array([
    [2.01, 2.99, 3.98 , 5.02],
    [20.78, -15.26, 19.38, -25.02],
    [3.08, 3.88, 4.15, 6.02]
    ])
```

[6]:
```
for i, x_i in enumerate(x_1):

    cov_norm = np.linalg.norm(cov)**1/2
    cov_inv = np.linalg.inv(cov)
    x_mu = x_i - mean
    print(f"p(x{i+1}): = ", f"{(1/ ( (2*np.pi)**2 * cov_norm )) * np.exp(-0.5 *␣
 ↪(x_mu).dot(cov_inv).dot((x_mu).T))[0][0][0]:.4f}")
```

```
p(x1): =  0.0051
p(x2): =  0.0000
p(x3): =  0.0030
```

Non Parametric Approach

We will use the Krnel Density Method to compute a gaussian kernel density estimate with an h of 1.
Source

[7]:
```
def parzen_window_est(x_samples, h, center):
    '''
    Implementation of the Parzen-window estimation for hypercubes.

    Keyword arguments:
        x_samples: A 'n x d'-dimensional numpy array, where each sample
            is stored in a separate row.
        h: The length of the hypercube.
        center: The coordinate center of the hypercube
```

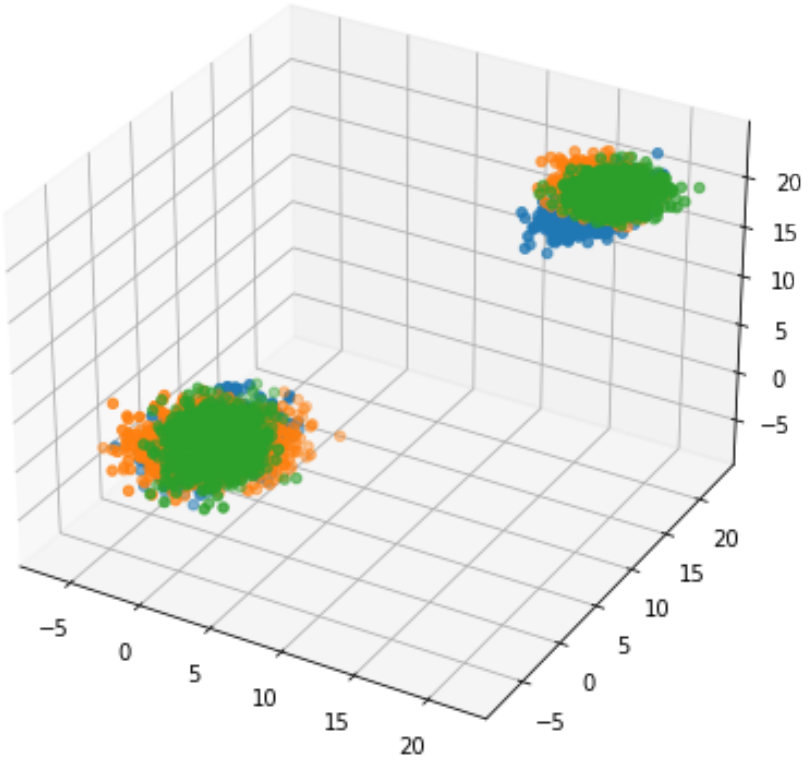
```
    Returns the probability density for observing k samples inside the hypercube.

    '''
    dimensions = x_samples.shape[1]

    assert (len(center) == dimensions), 'Number of center coordinates have to␣
 ↪match sample dimensions'
    k = 0
    for x in x_samples:
        is_inside = 1
        for axis,center_point in zip(x, center):
            if np.abs(axis-center_point) > (h/2):
                is_inside = 0
        k += is_inside
    return f"{(k / len(x_samples)) / (h**dimensions):.4f}"

for i, x_i in enumerate(x_1):
    # print('p(x) =', parzen_window_est(X1, h=1))
    print(f"p(x{i+1}) = ", parzen_window_est(X1, 1, x_i))
```

```
p(x1) =  0.0060
p(x2) =  0.0000
p(x3) =  0.0020
```

8.b Based on $X2$, estimate the values of $p_2(x)$ at the following points:

$x_1 = (0.05, 0.15, 0.12, 0.08), x_2 = (7.18, 7.98, 9.12, 9.94),$
$x_3 = (3.48, 4.01, 4.55, 4.96), x_4 = (20.78, 15.26, 19.38, 25.02)$

[8]:
```
fig = plt.figure(figsize = (10, 7))
ax = plt.axes(projection ="3d")


ax.scatter3D(X2[:,0], X2[:,1], X2[:,2])
ax.scatter3D(X2[:,0], X2[:,2], X2[:,3])
_ = ax.scatter3D(X2[:,1], X2[:,2], X2[:,3])
```



[9]:
```
# Parametric approach
gm_2 = GaussianMixture(n_components=2, random_state=0).fit(X2)
mean = gm_2.means_
cov = gm_2.covariances_
mean
```

[9]:
```
array([[ 1.69885937e+01,  1.79929660e+01,  1.89933557e+01,
         1.99915999e+01],
       [ 2.80527613e-03,  3.24565079e-02, -1.33126366e-01,
         8.13765609e-02]])
```

[10]:
```
# define X2 to take values for the predictions
x_2 = np.array([
    [0.05,0.15,-0.12,-0.08],
    [7.18,7.98,9.12,9.94],
    [3.48,4.01,4.55,4.96],
    [20.78,-15.26,19.38,-25.02]
    ])
```

[11]:
```
# for each x calculate the corresponding p(x)
for i, x_i in enumerate(x_2):
    p1 = gm_2.predict_proba([x_2[0]])[0][0]
    p2 = gm_2.predict_proba([x_2[0]])[0][1]

    cov_norm_1, cov_norm_2 = np.linalg.norm(cov[0])**1/2, np.linalg.
 ↪norm(cov[1])**1/2
    cov_inv_1, cov_inv_2 = np.linalg.inv(cov[0]), np.linalg.inv(cov[1])
    x_mu_1, x_mu_2 = x_i - mean[0], x_i - mean[1]

    p_x_1 = 1/ ( (2*np.pi)**2 * cov_norm_1 ) * np.exp(-0.5 * (x_mu_1).
 ↪dot(cov_inv_1).dot((x_mu_1).T))
    p_x_2 = 1/ ( (2*np.pi)**2 * cov_norm_2 ) * np.exp(-0.5 * (x_mu_2).
 ↪dot(cov_inv_2).dot((x_mu_2).T))

    P1_x = p1 * p_x_1 / (p1 * p_x_1 + p2 *p_x_2)
    P2_x = p2 * p_x_2 / (p1 * p_x_1 + p2 *p_x_2)
    print(f"p(x{i+1}): ", f"{p1 * p_x_1 + p2 *p_x_2:.4f}")
```

```
p(x1):  0.0050
p(x2):  0.0000
p(x3):  0.0000
p(x4):  0.0000
```

Non Parametric Approach

We will use the Kernel Density Method to compute a gaussian kernel density estimate with an h of 1. Source

[12]:
```
for i, x_i in enumerate(x_2):
    # print('p(x) =', parzen_window_est(X1, h=1))
    print(f"p(x{i+1}) = ", parzen_window_est(X2, 1, x_i))
```

```
p(x1) =  0.0005
p(x2) =  0.0000
p(x3) =  0.0000
p(x4) =  0.0000
```

- We observe that in both cases we get similar results for both methods.

Looking at the data points and the means of the distributions, it looks like our code has worked as expected. In the first case, the first point is $x_1 = (2.01, 2.99, 3.98, 5.02)$ and the mean is $\mu = (1.88364427, 2.94930523, 3.94277602, 4.91643708)$ with the corresponding $p(x_1)$ being $p(x_1) = 0.0060$.

On the other hand the second point is $x_2 = (20.78, 15.26, 19.38, 25.02)$ with $p(x_2) = 0.0060$.

This makes sense as the further away from the calculated mean we are, the lower the value of the probability density function will be.