Problem Set 2
ECE 590 Fall 2019

Name: Ana Belen Barcenas J.

Problem 1: Conditional Multivariate Gaussian Distribution

Solution:

A) Given the multivariate Gaussian distribution in the problem, the PDF of x_2 given $x_1 = a$ is determined by:

$$f(x_2|x_1 = a) = (2\pi)^{-D/2} det(\Sigma_{x_2|x_1=a})^{-1/2} e^{-\frac{1}{2}(x-\mu_{x_2|x_1=a})^T \Sigma^{-1}_{x_2|x_1=a}(x-\mu_{x_2|x_1=a})}$$

where

$$\Sigma = \begin{pmatrix} \mathbf{I}_2 & 0.25 * \mathbf{I}_2 \\ 0.25 * \mathbf{I}_2 & \mathbf{I}_2 \end{pmatrix}$$

and

$$\mu_{x_2|x_1=a} = \Sigma_{x_2|x_1=a}\{\Lambda_{x_2x_2}\mu_{x_2} - \Lambda_{x_2x_1}(x_1 - \mu_{x_1})\}$$

$$= \mu_{x_2} - \Lambda^{-1}_{x_2x_2}\Lambda_{x_2x_1}(x_1 - \mu_{x_1})$$

$$= \mu_{x_2} + \Sigma_{x_2x_1}\Sigma^{-1}_{x_1x_1}(x_1 - \mu_{x_1})$$

B) The mean vector is given by

$$\mu_{x_2|x_1=a} = \mu_{x_2} + \Sigma_{x_2x_1}\Sigma^{-1}_{x_1x_1}(x_1 - \mu_{x_1})$$

$$= \mu_{x_2} + \Sigma_{x_2x_1}\Sigma^{-1}_{x_1x_1}((1,1) - (1,1))$$

$$= \mu_{x_2} = (1,2)$$

And the covariance matrix by:

$$\Sigma_{x_2|x_1=a} = \Sigma_{x_2x_2} - \Sigma_{x_2x_1}\Sigma^{-1}_{x_1x_1}\Sigma_{x_1x_2} = 1.1875\,\mathbf{I}_2$$

Problem 4: Maximum Likelihood Estimation of Marchenko-Pastur distribution

Solution:

The Marchenko-Pastur distribution is given by:

$$(c, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} \mu(x_i|c, \sigma^2) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \frac{\sqrt{(c_+ - x_i)(x_i - c_-)}}{cx_i}$$

for all $x_i \in [c_-, c_+]$

$$= \Big(\frac{1}{2\pi\sigma^2}\Big)^n \Big(\frac{1}{c}\Big)^n \prod_{i=1}^{n} \frac{\sqrt{(c_+ - x_i)(x_i - c_-)}}{x_i}$$

for all $x_i \in [c_-, c_+]$

To make the computations easier, we are going to split some terms:

$$= \Big(\frac{1}{2\pi\sigma^2}\Big)^n \Big(\frac{1}{c}\Big)^n \prod_{i=1}^{n} \frac{1}{x_i} \prod_{i=1}^{n} \sqrt{(c_+ - x_i)(x_i - c_-)}$$

for all $x_i \in [c_-, c_+]$

now, let's apply log to each one separately and further simplify the Marchenko-Pastur distribution:

$$\ell(c, \sigma^2; x) = \log((c, \sigma^2; x)) = -n\log(2\pi\sigma^2 c) - \sum_{i=1}^{n} \log(x_i) + \frac{1}{2}\sum_{i=1}^{n}(\log(c_+ - x_i) + \log(x_i - c_-))$$

Finally, let's take the derivative in terms of c and σ^2 and set them to zero. This will give us the equations to find the MLE of each parameter:

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{-n}{2\pi\sigma^2 c} 2\pi c - \frac{1}{2}\sum_{i=1}^{n} \frac{(1+\sqrt{c}^2)}{\sigma^2(1+\sqrt{c})^2 - x_i} - \frac{(1-\sqrt{c})^2)}{x_i - \sigma^2(1-\sqrt{c})^2} = 0$$

$$\frac{\partial \ell}{\partial c} = \frac{-n}{2\pi\sigma^2 c} 2\pi\sigma^2 + \frac{1}{2}\sum_{i=1}^{n} \frac{\sigma^2(1+\sqrt{c})\sqrt{c}}{\sigma^2(1+\sqrt{c})^2 - x_i} - \frac{\sigma^2(1-\sqrt{c})\sqrt{c})}{x_i - \sigma^2(1-\sqrt{c})^2} = 0$$

Problem 5: Minimizing Minkowski Loss

Solution:

A) Loss function provided:

$$L(t, y(x)) = |t - y(x)|$$

To minimize the expected value of the loss function above, we have to compute the derivative of it with respect to $y(x)$. the expected value is given by:

$$\mathbb{E}(L) = \int \left[\int L(t, y(x)) p(x,t) dt \right] dx$$

Due to the law of total probability, we know that $p(x,t) = p(t|x)p(x)$, thus:

$$\mathbb{E}(L) = \int \left[\int L(t, y(x)) p(t|x) dt \right] p(x) dx$$

we can take $p(x)$ out of the first integral since it is fixed and could be considered as a constant.

Now, let's substitute the loss function provided:

$$\mathbb{E}(L) = \int \left[\int |t - y(x)| p(t|x) dt \right] p(x) dx$$

We can dismiss the outer integral since it is not dependant on $y(x)$ which is what we are going to minimize. Now, let's split the remaining integral based on the values t can take (integral limits) and then differentiate with respect to $y(x)$ and set it to zero:

$$\mathbb{E}(L) = \left[\int_{t=y(x)}^{\infty} (t - y(x)) p(x|t) dt \right] + \left[\int_{t=-\infty}^{y(x)} (y(x) - t) p(x|t) dt \right]$$

as we can see, $t * p(x|t)$ does not depend on y(x) what makes the derivative of those terms equal to zero and we can dismiss them

$$\mathbb{E}(L) = -\int_{t=y(x)}^{\infty} y(x) p(x|t) dt + \int_{t=-\infty}^{y(x)} y(x) p(x|t) dt$$

$$\frac{\partial \mathbb{E}(L)}{\partial y(x)} = -\int_{t=y(x)}^{\infty} p(x|t) dt + \int_{t=-\infty}^{y(x)} p(x|t) dt = 0$$

$$\mathbb{P}\Big(t > y(x) | x\Big) = \mathbb{P}\Big(t < y(x) | x\Big)$$

So we can confirm that the median of the conditional distribution minimizes $\mathbb{E}(L)$.

B) Loss function provided:

$$L(t, y(x)) = \begin{cases} 0 & |t - y(x)| \leq \delta \\ 1 & |t - y(x)| > \delta \end{cases}, \delta > 0$$

To minimize the expected value of the loss function above, we have to compute the derivative of it with respect to $y(x)$. the expected value is given by:

$$\mathbb{E}\big[L\big] = \iint_{|t-y(x)|>\delta} p(x,t)\, dx\, dt$$

as stated in problem A) above, $p(x,t) = p(t|x)p(x)$ and $p(x)$ does not depend on $y(x)$. We will dismiss $p(x)$ and differentiate the resulting integral:

$$\frac{\partial \mathbb{E}(L)}{\partial y(x)} = \int_{-\infty}^{y(\mathbf{x})-\delta} p(t|\mathbf{x})\, dt + \int_{\delta - y(\mathbf{x})}^{\infty} p(t|\mathbf{x})\, dt$$

$$= 1 - \int_{y(\mathbf{x})-\delta}^{y(\mathbf{x})+\delta} p(t|\mathbf{x})\, dt$$

Showing that the $\mathbb{E}(L)$ is minimized by the mode of the conditional distribution.