IITM-CS5691 : Pattern Recognition and Machine Learning — Release Date: August 30, 2023
Assignment 1 — Due Date : Wednesday, 27 September 2023, 11:59 PM

Roll No: CS23E001 — Name: Shuvrajeet Das
Collaborators (if any):
References/sources (if any):

- Use LaTeX to write-up your solutions (in the solution blocks of the source LaTeX file of this assignment), and submit the resulting pdf files (one per question) at Crowdmark by the due date. (Note: **No late submissions** will be allowed, other than one-day late submission with 10% penalty or four-day late submission with 30% penalty! Instructions to join Crowdmark and submit your solution to each question within Crowdmark **TBA** later).
- For the programming question, please submit your code (rollno.ipynb file and rollno.py file in rollno.zip) directly in moodle, but provide your results/answers (including Jupyter notebook **with output**) in the pdf file you upload to Crowdmark.
- Collaboration is encouraged, but all write-ups must be done individually and independently, and mention your collaborator(s) if any. Same rules apply for codes written for any programming assignments (i.e., write your own code; we will run plagiarism checks on codes).
- If you have referred a book or any other online material or LLMs (Large Language Models like ChatGPT) for obtaining a solution, please cite the source. Again don't copy the source *as is* - you may use the source to understand the solution, but write-up the solution in your own words (this also means that you cannot copy-paste the solution from LLMs!). Please be advised that *the lesser your reliance on online materials or LLMs* for answering the questions, *the more your understanding* of the concepts will be and *the more prepared you will be for the course exams*.
- Points will be awarded based on how clear, concise and rigorous your solutions are, and how correct your answer is. The weightage of this assignment is 12% towards the overall course grade.

1. (8 points) [GETTING YOUR BASICS RIGHT!]

   (a) (5 points) Let a random vector X follow a bivariate Gaussian distribution with mean $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, i.e., $X \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$. Then, use the pdf (probability density function) of X to:
   Find the distribution of (i) $X_2|X_1 = x_1$ and (ii) $X_1|X_2 = x_2$, and use them to (iii) find the permissible values of $a$, $b$, $c$, and $d$.
   (Hint: You can use the same approach of "completing the squares" seen in class).

(b) (2 points) Consider the function $f(x) = x_1^2 + x_2^2 + x_1x_2$, and a point $v = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$. Find the linear approximation of f around $v$ (i.e., $L_v[f](y)$), and show that the graph of this approximation is a hyperplane in $\mathbb{R}^3$.

(c) (1 point) Which of these statements are true about two random variables X and Y defined on the same probability space?

(i) If $X, Y$ are independent, then $X, Y$ are uncorrelated ($Cov(X, Y) = 0$).

(ii) If $X, Y$ are uncorrelated, then $X, Y$ are independent.

(iii) If $X, Y$ are uncorrelated and follow a bivariate normal distribution, then $X, Y$ are independent.

(iv) None of the above.

**Solution:** The solution of question(a)

Given the conditions, $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

from this we can get, $\Sigma^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ and $|\Sigma| = (ad - bc)$

The multivariate Gaussian distribution can defined as:-

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Putting all the necessary values we get:

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}} \exp\left(-\frac{1}{2(ad-bc)} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix})\right)$$

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}} \exp\left(-\frac{1}{2(ad-bc)} \begin{bmatrix} x_1d - x_2c & -x_1b + x_2a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$$

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}} \exp\left(-\frac{1}{2(ad-bc)}(x_1^2d - x_1x_2c - x_1x_2b + x_2^2a)\right)$$

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}} \exp\left(-\frac{1}{2(ad-bc)}(x_1^2d - x_1x_2(c+b) + x_2^2a)\right)$$

For question (1)$P(X_2|X_1 = x_1)$ is

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}} \exp\left(-\frac{a}{2(ad-bc)}\left(\frac{x_1^2d}{a} - 2x_1x_2\left(\frac{c+b}{2a}\right) + x_2^2\right)\right)$$

Let's solve the inner part in the exponent,

$$= \frac{-a}{2(ad-bc)}\left(\frac{x_1^2 d}{a} - 2x_1x_2\left(\frac{c+b}{2a}\right) + x_2^2\right)$$
$$= \frac{-a}{2(ad-bc)}\left(\frac{x_1^2 d}{a} - \left(\frac{c+b}{2a}\right)^2 x_1^2 + \left(\frac{c+b}{2a}\right)^2 x_1^2 - 2x_1x_2\left(\frac{c+b}{2a}\right) + x_2^2\right)$$
$$= \frac{-a}{2(ad-bc)}\left(\frac{x_1^2 d}{a} - \left(\frac{c+b}{2a}\right)^2 x_1^2 + \left(\left(\frac{c+b}{2a}\right)x_1 - x_2\right)^2\right)$$
$$= \frac{-a}{2(ad-bc)}\left(\left(\frac{4ad-(c+b)^2}{4a^2}\right)x_1^2 + \left(\left(\frac{c+b}{2a}\right)x_1 - x_2\right)^2\right)$$
$$= \frac{-a}{2(ad-bc)}\left(\frac{4ad-(c+b)^2}{4a^2}\right)x_1^2 + \frac{-a}{2(ad-bc)}\left(\left(\frac{c+b}{2a}\right)x_1 - x_2\right)^2$$

Simply filling the equation again we get, a product of two quantities,

$$\frac{1}{(2\pi)^{1/2}(ad-bc)^{1/2}}\exp\left(\frac{-a}{2(ad-bc)}\left(\left(\frac{c+b}{2a}\right)x_1 - x_2\right)^2\right)$$

and

$$\frac{1}{(2\pi)^{1/2}}\exp\left(\frac{-a}{2(ad-bc)}\left(\frac{4ad-(c+b)^2}{4a^2}\right)x_1^2\right)$$

For question (2)$P(X_1|X_2 = x_2)$ is

$$= \frac{1}{(2\pi)(ad-bc)^{1/2}}\exp\left(-\frac{d}{2(ad-bc)}\left(x_1^2 - 2x_1x_2\left(\frac{c+b}{2d}\right) + \frac{x_2^2 a}{d}\right)\right)$$

Let's solve the inner part in the exponent,

$$= -\frac{d}{2(ad-bc)}\left(x_1^2 - 2x_1x_2\left(\frac{c+b}{2d}\right) + \frac{x_2^2 a}{d}\right)$$
$$= \frac{-d}{2(ad-bc)}\left(x_1^2 - 2x_1x_2\left(\frac{c+b}{2d}\right) + \left(\frac{c+b}{2d}\right)^2 x_2^2 - \left(\frac{c+b}{2d}\right)^2 x_2^2 + \frac{x_2^2 a}{d}\right)$$
$$= \frac{-d}{2(ad-bc)}\left(\left(x_1 - \left(\frac{c+b}{2d}\right)x_2\right)^2 - \left(\frac{c+b}{2d}\right)^2 x_2^2 + \frac{x_2^2 a}{d}\right)$$
$$= \frac{-d}{2(ad-bc)}\left(\left(x_1 - \left(\frac{c+b}{2d}\right)x_2\right)^2 + \left(\frac{4ad-(c+b)^2}{4d^2}\right)x_2^2\right)$$
$$= \frac{-d}{2(ad-bc)}\left(x_1 - \left(\frac{c+b}{2d}\right)x_2\right)^2 + \frac{-d}{2(ad-bc)}\left(\frac{4ad-(c+b)^2}{4d^2}\right)x_2^2$$

Simply filling the equation again we get, a product of two quantities,

$$\frac{1}{(2\pi)^{1/2}(ad-bc)^{1/2}} \exp\left(\frac{-d}{2(ad-bc)}\left(x_1-\left(\frac{c+b}{2d}\right)x_2\right)^2\right)$$

and

$$\frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-1}{2(ad-bc)}\left(\frac{4ad-(c+b)^2}{4d^2}\right)x_2^2\right)$$

we can say that $d > 0$, $ad - bc > 0$, $a > 0$ also $a = d = 1$ filling up the values we get,

$$P(X_1|X_2) = \frac{1}{(2\pi)^{1/2}(1-bc)^{1/2}} \exp\left(\frac{-1}{2(1-bc)}\left(x_1-\left(\frac{c+b}{2}\right)x_2\right)^2\right)$$

$$P(X_2) = \frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-1}{2(1-bc)}\left(\frac{4-(c+b)^2}{4}\right)x_2^2\right)$$

$$P(X_2|X_1) = \frac{1}{(2\pi)^{1/2}(1-bc)^{1/2}} \exp\left(\frac{-1}{2(1-bc)}\left(\left(\frac{c+b}{2}\right)x_1-x_2\right)^2\right)$$

$$P(X_2) = \frac{1}{(2\pi)^{1/2}} \exp\left(\frac{-1}{2(1-bc)}\left(\frac{4-(c+b)^2}{4}\right)x_1^2\right)$$

we also get $1 - bc > 0$

The solution of question (b)

we know from the taylor series expnasion,

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

The linear approximation of a function at a given point is its first-order Taylor series expansion around that point. To find the linear approximation of the function $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2$, we need to choose a point (a, b) around which we want to approximate the function. Let's choose the point (a, b).

The linear approximation of $f(x_1, x_2)$ at the point (a, b) is given by:

$$L(x_1, x2) = f(a,b) + \frac{\partial f(a,b)}{\partial x_1}(x_1-a) + \frac{\partial f(a,b)}{\partial x_2}(x_2-b)$$

To find the linear approximation, we need to calculate the partial derivatives of $f(x_1, x_2)$ with respect to x and y. Here are the derivatives:

$$\frac{\partial f(x_1,x_2)}{\partial x_1} = 2x_1 + x_2 \quad \frac{\partial f(x_1,x_2)}{\partial x_2} = 2x_2 + x_1$$

Now, we can plug these derivatives into the linear approximation formula:

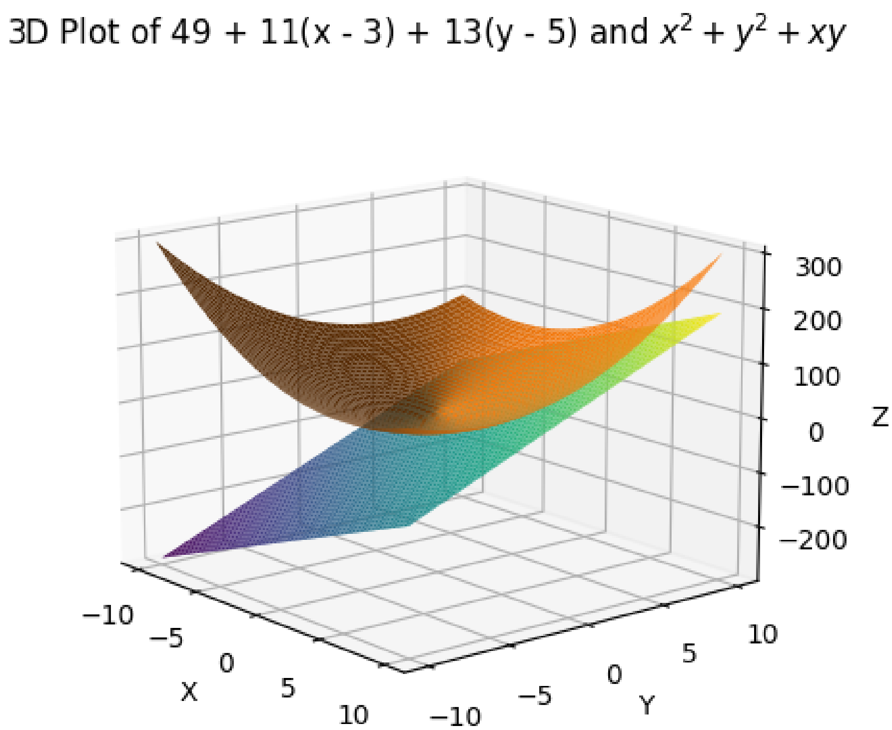
$$L(x_1, x_2) = (a^2 + b^2 + ab) + (2a + b)(x_1 - a) + (2b + a)(x_2 - b)$$

So, the linear approximation of $f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2$ around the point (a, b) is:

$$L(x_1, x_2) = a^2 + b^2 + ab + (2a + b)(x_1 - a) + (2b + a)(x_2 - b)$$

Filling up all the necessary values we get

$$L(x_1, x_2) = 49 + 11(x - 3) + 13(y - 5)$$




The solution of question (c)

The statement no (i) and (iii) are correct.