SIG 787 - Mathematics for AI
Assignment 3

Trimester 1, 2023

Important notes:

- Your submission can be handwritten but it must be legible. If your submission is not legible, it will not be marked and will result in a zero mark. A proper way of presenting your solutions is part of the assessment.
- Please follow the order of questions in your submission.
- All steps (workings) to arrive at the answer must be clearly shown. No marks will be awarded for answers without workings.
- Generally, you need to keep your answers in the form of quotients and surds (e.g. $\frac{2}{3}$ and $\sqrt{3}$). Rarely, you may convert your solutions into decimals for plotting or comparing purposes. However, you need to show the final answer in terms of quotients and surds before converting them into decimals.
- Only (scanned) electronic submission would be accepted via the unit site (DeakinSync).
- Your submission must be in ONE pdf file. Multiple files and/or in different file format, e.g. .jpg, will NOT be accepted. It is your responsibility to ensure your file is not corrupted and can be read by a standard pdf viewer. Failure to comply will result in a zero mark.

Question 1) Graph Theory: Consider the following incidence matrix of a graph $G = (V, E)$ with $V = \{a, b, c, d\}$ and $\{e_1, e_2, e_3, e_4, e_5, e_6\}$

$$M = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 & 0 & -1 \end{pmatrix} \end{array}$$

Based on the information you obtain from the incidence matrix M, answer these questions:

(a) What type of graph does M represent?

(b) Find the adjacency matrix A for this graph.

(c) Draw the graph.

(d) How many paths of length 2 are there between nodes b and c (without direct counting)?

(e) In terms of connectivity of the graph, what is your interpretation of $\mathrm{tr}(A^2)$? [The question is not about the value.] [2+2+2+2+2=10 marks]

Question 2) Probability, Bayes' Theorem: A data scientist is interested in studying the relationship between having a social media account and obesity. He collected some data represented in the following table.

Individual	**Having a social media account**	**Obesity**
Individual1	Yes	Yes
Individual2	Yes	Yes
Individual3	Yes	Yes
Individual4	No	No
Individual5	No	Yes

(a) Find the probability distributions of having a social media account and being obese.

(b) Find the joint probability distribution of having a social media account and being obese.

(c) Are these two distributions independent?

(d) The data scientist is interested in whether having a social media account helps gain information about obesity. The concept of mutual information quantifies how one feature is related to another future. For two random variables X and Y, the mutual information is defined as follows.

$$\text{mutual information} = \sum_x \sum_y p_{X,Y}(x,y) \ln\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right).$$

Suppose the joint probability distribution of having a social media account and obesity is shown by $p_{X,Y}(x,y)$ and the marginal distributions of having a social media account and obesity are $p_X(x)$ and $p_Y(y)$ respectively. Compute the mutual information between having a social media account and obesity. [You can use a calculator to find $\ln(x)$.]

(e) Theoretically, what is the mutual information of two independent random variables?

[5+5+5+5+5= 25 marks]

Question 3) Probability, Distributions: Let X be a discrete random variable that takes values in $\{-2,-1,0,1,2\}$ with equal probability. Also, Y is another discrete random variable defined as $Y = |X|$.

(a) Construct the joint probability distribution table.

(b) Are X and Y independent? Justify.

(c) Find $\text{Corr}(X,Y)$.

(d) Based on your answer to part (b), can you explain the result in part (c)?

(e) Although you have solved similar questions in questions 2 and 3, can you explain the fundamental difference between these two questions?

[5+5+5+5+5= 25 marks]

Multivariate functions and Optimisation. The idea of local and global minimum and maximum exists in studying functions with several variables. For a function of two variables $z = f(x, y)$, if f has a local maximum or minimum at a point (a, b), and the first order partial derivatives of f exist there, then $f_x(a, b) = 0$ and $f_y(a, b) = 0$. You can imagine that if $f_x(a, b) = 0$ and $f_y(a, b) = 0$, then there will be a flat tangent plane, similar to a horizontal tangent line in single variable functions.

Formally, a point (a, b) is a critical point (or stationary point) of $z = f(x, y)$ either if $f_x(a, b) = 0$ and $f_y(a, b) = 0$, or if one of these partial derivatives does not exist. Also, we need to be careful that at a critical point, a multivariate function could have a local maximum, a local minimum or neither.

Second derivative test for a local max or min for functions with 2 variables: Consider $z = f(x, y)$. We need to be able to determine whether a function has an extreme value at a critical point. For a critical point (a, b), let

$$D = D(a, b) = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$$

be the discriminant.

- If $D > 0$ and $f_{xx}(a, b) > 0$, then f has a local minimum at (a, b).
- If $D > 0$ and $f_{xx}(a, b) < 0$, then f has a local maximum at (a, b).
- If $D < 0$, then f does not have a local max or a local min at (a, b). It has a saddle point there.
- If $D = 0$, then f has a degenerate critical point at (a, b), and the test gives no information. We need to use other mathematical techniques to check the situation.

To find a local minimum, a local maximum, or a saddle point of multivariate functions with more than 2 variables, we need to compute its Hessian matrix at all the critical points. Consider the multivariate function $w = f(x, y, z)$. You see that the number of independent variables is more than 2. The Hessian of this matrix is defined as

$$Hf(x, y, z) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

Generally, $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$, $\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x}$, and $\frac{\partial^2 f}{\partial z \partial y} = \frac{\partial^2 f}{\partial y \partial z}$, hence, the Hessian matrix is usually symmetric. Also, note that the Hessian is a function of x, y, and z. The critical points of this function are either $\nabla f = \overrightarrow{\mathbf{0}}$, or when at least one of the partial derivatives does not exist. To find a local minimum, a local maximum, or a saddle point, we plug the critical points in the Hessian matrix, compute its eigenvalues, and we make a decision based on signs of the eigenvalues, as described in the second derivative test.

Second derivative test: Find all critical points, plug them in the Hessian matrix, and compute their eigenvalues.

- If **all** eigenvalues are strictly positive, then the critical point is a local minimum.
- If **all** the eigenvalues are strictly negative, the critical point is a local maximum.
- If eigenvalues do not have the same sign and **all** are non-zero, then the critical point is a saddle point.
- If any of the eigenvalues is zero, we say that the critical point is degenerate. No conclusion can be drawn from this case and must use another method to determine whether or not it is an extreme point of f.

It is important to note that by this extension, you can find the local maximum, local minimum, or saddle points of a function with any number of variables. In its most general form, to find critical points of a multivariate function with n variables $y = f(x_1, x_2, \ldots, x_n)$, you need to solve the system $\nabla f = \overrightarrow{\mathbf{0}}$, or find the points that any of the partial derivatives do not exist. Then, you need to form the Hessian matrix at the critical points, find eigenvalues, and make a decision based on the signs of the eigenvalues.

For a critical point $\mathbf{a} = (a_1, \ldots, a_n)$, $y = f(x_1, x_2, \ldots, x_n)$ the Hessian matrix is evaluated as follows:

$$Hf(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & f_{x_1x_3}(\mathbf{a}) & \ldots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & f_{x_2x_3}(\mathbf{a}) & \ldots & f_{x_2x_n}(\mathbf{a}) \\ f_{x_3x_1}(\mathbf{a}) & f_{x_3x_2}(\mathbf{a}) & f_{x_3x_3}(\mathbf{a}) & \ldots & f_{x_3x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots & \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & f_{x_nx_3}(\mathbf{a}) & \ldots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}$$

There is an alternative but equivalent second derivative test to classify critical points. From the Hessian, calculate the sequence of principal minors of $Hf(\mathbf{a})$. This is the sequence of the determinants of the upper leftmost square submatrices of $Hf(\mathbf{a})$. More explicitly, this is the sequence $d_1, d_2, \ldots, d_n$, where $d_k = \det(H_k)$, and H_k is the upper leftmost $k \times k$ submatrix of $Hf(\mathbf{a})$. That is,

$$d_1 = f_{x_1x_1}(\mathbf{a})$$

$$d_2 = \begin{vmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) \end{vmatrix}$$

$$d_3 = \begin{vmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & f_{x_1x_3}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & f_{x_2x_3}(\mathbf{a}) \\ f_{x_3x_1}(\mathbf{a}) & f_{x_3x_2}(\mathbf{a}) & f_{x_3x_3}(\mathbf{a}) \end{vmatrix}$$

$$\vdots$$

$$d_n = \det(Hf(\mathbf{a}))$$

The numerical test is as follows: Assume that $\det(Hf(\mathbf{a})) \neq 0$.

1. If $d_k > 0$ for $k = 1, 2, \ldots, n$, then f has a local minimum at $\mathbf{a}$.
2. If $d_k < 0$ for k odd, and $d_k > 0$ for k even, then f has a local maximum at $\mathbf{a}$.
3. If neither case 1 nor case 2 holds, then f has a saddle point at $\mathbf{a}$.

In the event that $\det(Hf(\mathbf{a})) = 0$, we say that the critical point $\mathbf{a}$ is degenerate and must use another method to determine whether or not it is the site of an extremum of f.

Constrained Optimisation

So far what we had was unconstrained optimisation. In other words, we wanted to find a local maximum or minimum, or a saddle point of a multivariate function without any restriction or constraints. However, sometimes we want to find a maximum or a minimum of $f(x, y)$ subject to a constraint $g(x, y) = k$, where k is a constant. To solve this problem, we use the Lagrange Multiplier technique. To find maximum and minimum values of $f(x, y)$ subject to the constraint $g(x, y) = k$ (assuming that these max and min values exist and $\nabla g \neq 0$ on $g(x, y) = k$):

(a) Find all values of x, y, and λ such that

$$\nabla f(x, y) = \lambda \nabla g(x, y) \text{ and}$$

$$g(x, y) = k$$

(b) Evaluate f at all the points (x, y) that result from step (a). The largest of these values is the global maximum value and the smallest is the global minimum value of f.

[**Note:** Please pay attention to the way the λ notation is used here. It is not an eigenvalue. Here, it is a nonzero multiplier – we are not interested in its value, but it will facilitate the finding of the critical points.]

This technique could be applied for functions having more than two variables. Imagine, you want to find the maximum and minimum of $w = f(x, y, z)$ subject to constraint $g(x, y, z) = 0$. Then you need to solve the following system:

$$\begin{cases} \nabla f(x, y, z) = \lambda \nabla g(x, y, z) \\ g(x, y, z) = 0 \end{cases}$$

This system has four equations and 4 unknowns. However, generally, the equations are not linear. Sometimes, you may use some mathematical tricks to find the solution to the system of equations.

A further extension is when there are two constraints (you can see that we can easily generalise the idea to many constraints). To find the maximum and minimum of a function $w = f(x, y, z)$ subject to two constraints $g(x, y, z) = 0$ and $h(x, y, z) = 0$, we need to introduce two Lagrange multipliers λ and μ to obtain the following system of equations:

$$\begin{cases} \nabla f(x, y, z) = \lambda \nabla g(x, y, z) + \mu \nabla h(x, y, z) \\ g(x, y, z) = 0 \\ h(x, y, z) = 0 \end{cases}$$

To summarise, you need to be familiar with the following optimisation problems, and how we can solve them. Also, you need to be extremely careful about using local or global adjectives in your communications.

- Optimisation problems

  A. Unconstrained Optimisation

I. Single-variable function: (covered in assignment 1)

II. Multivariate functions

i. Multivariate with two variables (use the discriminant)

ii. Multivariate with more than two variables (use eigenvalues)

B. Constraint Optimisation

I. Global max and min of a single variable function in an interval (covered in assignment 1)

II. Global max and min of multivariate functions with some constraints (Lagrange multipliers)

Question 4) Now, let's solve some optimisation problems.

(a) For a function $f(x, y) = x^2 e^{-xy}$ find all directions $\vec{\mathbf{u}} = \begin{bmatrix} a \\ b \end{bmatrix}$ so that the directional derivative at the point $(1, 0)$ in the direction of $\vec{\mathbf{u}}$ is 1, $D_{\vec{\mathbf{u}}} f(1, 0) = 1$.

(b) For the multivariate function

$$f(x, y, z) = x^2 + y^2 + z^2 + yz$$

(i) Find the stationary point(s) of this function.

(ii) Find the Hessian matrix.

(iii) Find the eigenvalues and eigenvectors of the Hessian matrix at the stationary point(s).

(iv) Classify the stationary point(s).

(c) Find the extreme values of $f(x, y) = e^{xy^2}$ subject to $x + 2y = 3$ and $x - y = 0$.

[10+[5+4+7+4]+10 = 40 marks]