THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

This exam paper must not be removed from the venue

Venue	____________________
Seat Number	________
Student Number	|__|__|__|__|__|__|__|__|
Family Name	_____________________
First Name	_____________________

School of Mathematics & Physics

EXAMINATION

Semester Two Final Examinations, 2019

MATH7502 Mathematics for Data Science 2

This paper is for St Lucia Campus students.

Examination Duration: 180 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a Central Examination

This is a Closed Book Examination - specified materials permitted

During reading time - write only on the rough paper provided

This examination paper will be released to the Library

Materials Permitted In The Exam Venue:

(No electronic aids are permitted e.g. laptops, phones)

Calculators - Any calculator permitted - unrestricted

One A4 sheet of handwritten or typed notes double sided is permitted

Materials To Be Supplied To Students:

None

Instructions To Students:

Additional exam materials (eg. answer booklets, rough paper) will be provided upon request.

For Examiner Use Only

Question	Mark
1a	
1b	
2a	
2b	
3a	
3b	
4a	
4b	
5a	
5b	

Total ________

Instructions:

There are 5 questions, with 20 points each. Each question has two items. Item (a) worth 14 points and item (b) worth 6 points.

Answer all questions on this exam paper.

Use the other side of the paper if additional space is required.

Question 1

(a) You may use the singular value decomposition (SVD) to answer this question. Consider a 1000×50 data matrix A, summarizing data of 1000 individuals and 50 features with $\text{rank}(A) = 49$. Denote the eigenvalues of AA^T by $\lambda_1, \lambda_2, \ldots$ with $\lambda_i \geq \lambda_j$ for $j > i$. What is λ_{50}? Explain your answer.

(b) Assume now that $\lambda_1 = \lambda_2 = \ldots = \lambda_{49} = 4$. Determine,

$$\sqrt{\sum_{i=1}^{1000} \sum_{j=1}^{50} A_{ij}^2}.$$

Question 2

(a) Assume you are presented with pairs of data points $(x_1, y_1), \ldots, (x_n, y_n)$ where $x_i \neq x_j$ for $i \neq j$. To describe this data, you wish to fit a model,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2,$$

by selecting $\beta = [\beta_0, \beta_1, \beta_2]^T$ that will minimize,

$$\sqrt{\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2\right)^2}.$$

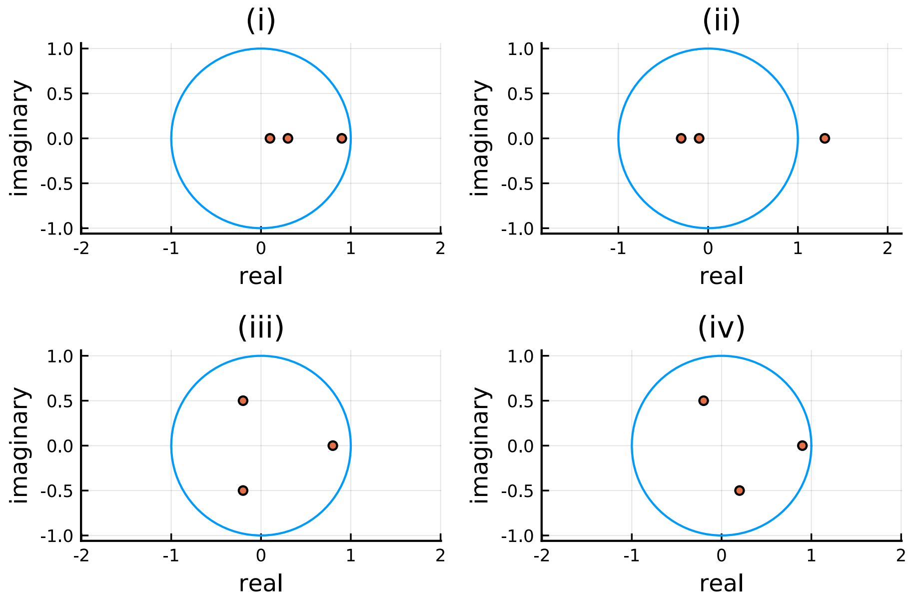
You do this by setting $\hat{\beta} = (A^T A)^{-1} A^T v$ where $A \in \mathbb{R}^{n\times 3}$ and $v \in \mathbb{R}^n$. Determine A and v.

(b) Assume now that you wish to find $\hat{\beta}$ using gradient descent. For this you compute the gradient descent step with learning rate $\eta > 0$, via,

$$\beta_{t+1} = \beta_t - \eta 2A^T(A\beta_t - v) = (I - 2\eta A^T A)\beta_t + 2\eta A^T v = G(\eta)\beta_t + 2\eta A^T v,$$

where the matrix $G(\eta) = I - 2\eta A^T A$ depends on the learning rate. You set the learning rate at a low enough value, η^* such that $\beta_t \to \hat{\beta}$ (gradient descent converges).

Let the eigenvalues of $G(\eta^*)$ be λ_1, λ_2 and λ_3. Below are four alternative plots of λ_1, λ_2 and λ_3 on the complex plane. For each of them, determine if it is possible or not and if not, explain why.



(i) Possible / Impossible:

(ii) Possible / Impossible:

(iii) Possible / Impossible:

(iv) Possible / Impossible:

Question 3

(a) Consider the vectors $v_1 = [1,1,0,0]^T$, $v_2 = [1,1,1,1]^T$ and the 4×2 matrix $A = [v_1 \; v_2]$. Determine a QR factorization of A having the form $A = QR$ where Q is a 4×2 matrix with orthonormal columns and R is an upper triangular matrix. Choose Q and R such that neither have negative entries. Throughout this question, avoid using decimal points, but rather use exact arithmetic.

(b) Let $b = [0\ 1\ 1\ 0]^T$. You now wish to find an approximate solution to $Ax = b$ in the sense that x minimizes $||Ax - b||$ ($||\cdot||$ is the L_2 norm). Use your QR factorization from the previous exercise to find x.

Question 4

(a) Consider a collection of N, n-vectors on which you execute the k-means algorithm. Denote the vectors by $x_1, \ldots, x_N$ and denote by G_j the set of indices of vectors in group j resulting from k-means. Thus for example if $G_3 = \{17, 21, 302\}$ then the vectors x_{17}, x_{21} and x_{302} make up group 3. Denote,

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_j, \qquad \text{for} \qquad j = 1, \ldots, k.$$

In general, k-means is only a heuristic and its solution does not always exactly minimize the clustering objective,

$$J = \sum_{j=1}^{k} \sum_{i \in G_j} ||x_i - z_j||^2.$$

However, sometimes it does. For part (a), assume that $n = 2$, $k = 2$ and $N = 6$ with,

$$x_1 = \begin{bmatrix} 9 \\ 9 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 11 \\ 11 \end{bmatrix}, \quad x_4 = \begin{bmatrix} -9 \\ -9 \end{bmatrix}, \quad x_5 = \begin{bmatrix} -10 \\ -10 \end{bmatrix}, \quad x_6 = \begin{bmatrix} -11 \\ -11 \end{bmatrix}.$$

Find G_1, G_2, z_1 and z_2 that minimize J.

(b) Returning to general n, N and k. Does it always hold that,

$$\sum_{i \in G_j} ||x_i - z_j||^2 \leq \sum_{i \in G_j} ||x_i - z||^2,$$

for any $j \in \{1, \ldots, k\}$ and any $z \in \mathbb{R}^n$? If so prove why, otherwise, explain why not or present a counter-example.

Question 5

(a) Consider the linear dynamical system, $x_{t+1} = Ax_t$ with,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Find the eigenvalues of A and corresponding eigenvectors where the second coordinate of each eigenvector is 1. To simplify notation, you may use the golden ratio $\psi = \frac{1}{2}(1+\sqrt{5})$ as well the fact that $1-\psi = \frac{1}{2}(1-\sqrt{5})$ and $\psi = 1/(\psi - 1)$.

(b) Assume now that $x_0 = [\psi \ 1]^T$. And denote the scalar q_t by,

$$q_t = [1 \ 0]x_t.$$

What is the smallest (integer) t for which $q_t \geq 1000$?

END OF EXAMINATION