Mathematical Foundations of Data Science
(COMS4055A)
Class Test 1

12 April 2022, 08h00–10h00, Flower Hall

Name: [redacted] Row: ______ Seat: ______ Signature: [redacted]

Student Number: [redacted] ID Number: [redacted]

For marking purposes only

Question 1	
Question 2	
Total	

Instructions

- Answer all questions in pen. **Do not write in pencil.**
- This test consists of 5 pages. Ensure that you are not missing any pages.
- This is a **closed-book** test: you may not consult any written material or notes.
- You are allocated 2 hours to complete this test.
- There are 2 questions and ~~80~~ 50 marks available.
- Ensure your cellphone is switched off.
- You may use a calculator during the test.

Question 1 Linear Algebra [40 Marks]

1. Consider the set $\mathcal{S} = \mathbb{R}\backslash\{-2\}$ with the operator $\circ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that: [6]

$$a \circ b = 2ab + 2a + 2b \text{ with } a, b \in \mathbb{R}\backslash\{-2\}$$

State the properties of an Abelian group and prove that $(\mathcal{S}, \circ)$ is an Abelian group.

2. Consider the set $\mathcal{S}$ of 3×3 matrices: [6]

$$\mathcal{S} = \{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times3} | x, y, z \in \mathbb{R}\}$$

Is $(\mathcal{G}, +)$ a group, where $+$ denotes element-wise addition? Is it Abelian?.

3. According to **Definition 2.10** of the course textbook, if $V = (\mathcal{V}, +, \cdot)$ is a vector space and we have a subset $\mathcal{U} \subseteq \mathcal{V}$ with $\mathcal{U} \neq \emptyset$ then $U = (\mathcal{U}, +, \cdot)$ is a vector subspace of V if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$ respectively. With this definition in mind, which of the following are subspaces of $\mathbb{R}^3$? Show your workings and give clear reasoning for why or why not.

   (a) $A = \{(\alpha, \alpha + \beta + \delta^3, \beta - \delta^2) | \alpha, \beta, \delta \in \mathbb{R}\}$ [3]

   (b) $B = \{(\mu^3, -\mu^4, 0) | \mu \in \mathbb{R}\}$ [3]

   (c) $C = \{(\gamma_1, \gamma_2, \gamma_3) \in \mathbb{R}^{+3} | \gamma_1 + \gamma_2 + \gamma_3 = 0\}$ (where $\mathbb{R}^+$ means the set of positive real values). [3]

4. A company has 4 products. Each product uses a combination of the same 3 resources. Product x_1 uses $(1, 3, 2)$ of each resource respectively. Likewise the resources used for x_2, x_3 and x_4 are $\cancel{(-1, 1, -1)}$, $(-2, 5, 0)$, $(-3, 1, -2)$ and $(1, 0, -1)$ respectively. In total the company has $(3, 6, 5)$ of each resource. Find the general solution for the number of products produced $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ which ensures the company uses all of the available resources. [7]

5. Consider two subspaces of $\mathbb{R}^4$.

$$U_1 = span[\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}] = span\left[\begin{bmatrix}1\\2\\-3\\0\end{bmatrix}, \begin{bmatrix}4\\-1\\-1\\1\end{bmatrix}, \begin{bmatrix}10\\-7\\3\\3\end{bmatrix}\right];$$

$$U_2 = span[\mathbf{v_4}, \mathbf{v_5}, \mathbf{v_6}, \mathbf{v_7}] = span\left[\begin{bmatrix}-1\\-2\\2\\1\end{bmatrix}, \begin{bmatrix}6\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\-6\\2\\1\end{bmatrix}, \begin{bmatrix}1\\0\\2\\0\end{bmatrix}\right]$$

(a) State the two properties of a basis for a vector space. [1]

(b) Determine a basis of U_1. [2]

(c) Determine a basis of U_2. [2]

(d) Determine a basis for the union $U_1 \cup U_2$. [2]

(e) Determine a basis for the intersection $U_1 \cap U_2$. [5]

Question 2 Analytic Geometry [40 Marks]

1. Compute the distance between

$$x = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \; y = \begin{bmatrix} -2 \\ -2 \\ 7 \end{bmatrix}$$

using:

(a)

$$< \mathbf{x}, \mathbf{y} >= x^T y$$

[2]

(b)

$$< \mathbf{x}, \mathbf{y} >= x^T \begin{bmatrix} 3 & 4 & 0 \\ 0 & 2 & -1 \\ 5 & 1 & 7 \end{bmatrix} y$$

[3]

2. Compute the angle between

$$x = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \; y = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

using:

(a)

$$< \mathbf{x}, \mathbf{y} >= x^T y$$

[2]

(b)

$$< \mathbf{x}, \mathbf{y} >= x^T \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} y$$

[3]

3. Consider a subspace $U \subseteq \mathbb{R}^3$ and $x \in \mathbb{R}^3$ as given by

$$U = span \left[\begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix} \right]$$

and

$$x = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$$

with the inner product:

$$< \mathbf{x}, \mathbf{y} >= x^T \begin{bmatrix} 2 & 3 & 0 \\ 0 & 2 & -1 \\ -1 & 0 & 2 \end{bmatrix} y$$

(a) Determine the orthogonal projection $\pi_U(x)$ of x onto U. [10]

(b) Determine the distance $\mathbf{d}(\mathbf{x}, \pi_{\mathbf{U}}(\mathbf{x}))$. [3]

4. Let V be a vector space and π an endomorphism of V. Prove that π is a projection if and only if $id_V - \pi$ is a projection, where id_V is the identity endomorphism on V. [6]

5. Using the Gram-Schmidt method, turn the basis $B = (b_1, b_2)$ of a two-dimensional subspace $U \subseteq \mathbb{R}^3$ into an Ortho-normal Basis $C = (c_1, c_2)$ of U, where [7]

$$b_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \; b_2 = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$$

$u_2 = b_2 - \frac{\langle u_1, b_2 \rangle}{\langle u_1, u_1 \rangle} b_1$

using the inner product:

$$< \mathbf{x}, \mathbf{y} >= x^T \begin{bmatrix} 3 & 0 & 1 \\ -2 & 0 & -1 \\ 0 & 2 & 0 \end{bmatrix} y$$

$u_2 = b_2 - \frac{\langle u_1, b_2 \rangle}{\langle u_1, u_2 \rangle} u_1$

6. You are given a set of data $X = \{(x_i, y_i) : i = 1...n\}$. On average when $x_i = 0, y_i \neq 0$ (in other words the data does not pass through the origin).

(a) Does this data exist on a vector or affine space (use the most specific definition possible according to the textbook definitions)? Justify your answer. [2]

(b) Assume there is a linear relationship between x_i and y_i. Based on your above answer would you include a bias parameter in a linear regression model? Justify your answer. [2]

$\pi_U(x) = B(B^T A B)^{-1} B^T A x$