

MTech in Data Science
IIT Palakkad

DS5003A : Data Engineering

0930-1230

Final Exam (22 Nov 2025)

Marks : 20

Instructions

1. Use Blue/ Black ink (red ink and pencils are not allowed). If your answer is not legible, you will not get any marks for that.
2. Do not use any language specific syntax or library routines.

Consider the following schema and answer the questions.

instructor(ID, name, salary)
student(ID, name, tot_cred)
course(course_ID, title, dept_name, credits)
department(dept_name, building, budget)
teaches(ID, course_ID, sec_ID, semester, year)
takes(ID, course_ID, semester, year, grade)
prereq(course_ID, prereq_ID)

branch(branch_name, branch_city, assets)
customer(customer_name, customer_street, customer_city)
loan(loan_number, branch_name, amount)
borrower(customer_name, loan_number)
account(account_number, branch_name, balance)
depositor(customer_name, account_number)
executive(employee_name, branch_name, customer_name)

1. Write a solution based on SQL statements to show the name of the instructors who teach courses, such that all students crediting those courses sit in the same building as of the instructor. (Hint. you may write multiple SQL statements.)
2. Write a single SQL statement to show the names of the customers for whom thier city of residence, city of account, and city of loan are same.
3. `select customer_name from account, customer where balance > 5000;`
Is there anything wrong in the above statement ? If yes, then write the correct statement with minimal change.
4. Write an algorithm on matrix multiplication. Make it as generally applicable as possible. Do not use any language specific syntax, and special functions. You should take care of required boundary conditions. Input will be two matrices, and output will be one matrix.

5. There is a matrix $X \in \mathcal{R}^{n \times m}$. There are 3 more matrices Q, K, V whose dimensions are not given. You have to write an algorithm to get another matrix $S \in \mathcal{R}^{n \times n}$, where S_{ij} is some score between x_i and x_j where x_i is the i^{th} row in X . The equation for S_{ij} is as follows:

$$\begin{aligned} a_i &= Qx_i, \quad b_i = Kx_i, \quad c_i = Vx_i \\ a_j &= Qx_j, \quad b_j = Kx_j, \quad c_j = Vx_j \\ S_{ij} &= (a_i^T b_j) c_j, \quad S_{ji} = (a_j^T b_i) c_i \end{aligned} \tag{1}$$

Note that, S need not be a symmetric matrix. Do not use any language specific syntax, and special functions. You should take care of required boundary conditions.

6. We want to build an image search algorithm. There is a collection of color images of various types like cars, vegetables, pets, bedrooms etc. Each image size can be different. Like keywords in text search, we have a set of visual keywords. A visual keyword can be like a small color image. You have to write an algorithm to search in the collection of images, which images contain a subpart which is similar to the visual keyword. Input to the algorithm will be a collection of images and a visual keyword. Output of the algorithm will be a score between 0 and 1 which denotes possibility of an image to contain the visual keyword, 0 being least possible, and 1 means most possible. Note that, size of a visual keyword is also unknown but is smaller than size of any image in the collection. Do not use any language specific syntax, and special functions. You can assume a function `loadImage` which will convert any image or visual keyword into a tensor.