

Data Mining: Practical Assignment #1

Due on Thu & Fri, April 6-7 2017, 10:15am-13:15 & 14:15am-17:15

Task 1

Get familiar with Matlab/Octave:

i.) How you generate the column vector $A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$?

ii.) Given a matrix $B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

Which of the following expressions would give the matrix $C = \begin{bmatrix} 2 & 3 \\ 5 & 6 \\ 8 & 9 \end{bmatrix}$?

1. `B(:, :)`

2. `B(:, 2)`

3. `B(:, 2:3)`

4. `B(2:3, :)`

iii.) Suppose you wish to generate a 3x1 vector D that contains the number 5 in every position.

Which of the following expressions will accomplish this task?

1. `eye(3)*5`

2. `ones(3)*5`

3. `ones(3,1)*5`

4. `fives(3,1)`

iv.) Which expression allows to create a new matrix E by appending the column vector D to the matrix B ?

v.) Suppose you wish to generate a 2x3 matrix F that contains only zeros.

Which of the following expressions will achieve this goal?

1. `zeros(2,3)`

2. `zeros(3,2)`

3. `[0 0 0; 0 0 0]`

4. `[0 0; 0 0; 0 0]`

vi.) Create a matrix M as follows: $M = \text{randi}([-5 \ 5], [6, \ 6])$.

How can you achieve the following matrix manipulation?

1. Get only the first row of M .

2. Get only the 5th column of M .

3. Get only the even rows of M.
4. Get only the odd columns of M.
5. Multiply the 5th column of M with 5.
6. Create a 3x3 matrix (M2) out of the bottom right corner of M.
7. Get the identity matrix of M2 (by multiplying M2 with its inverse).
8. Set all positive numbers of M to 0.
9. Set all even rows of M to 1.

Task 2

Data Sets: Load the remaining data files from the *data1.zip* into your Matlab workspace.

[Hint: Check the *ReadMe.txt* for information on the data sets.]

Use the plot list provided in the Matlab workspace panel. Decide, which data visualization method would be appropriate for which data set. We propose the following methods:

1. Scatter plot
2. Bar chart
3. Boxplot
4. Time series
5. Pie chart

Discuss your choices and your results of the plots.

@home Task

To prepare the next tutorial, your homework will be to learn about the following topics:

- χ^2 -test
- Covariance and correlation
- Principal Component Analysis (PCA), eigenvalues and eigenvectors

Recommended literature:

1. Kantardzic, M. Data Mining: Concepts, Models, Methods, and Algorithms (2nd Ed.). Wiley-IEEE Press, 2011.
2. Han J. & Kamber, M. Data mining: Concepts and techniques. Elsevier/Morgan Kaufmann, Amsterdam, 2006.

Both books are available in the Informatics library.

You are further encouraged to seek appropriate literature. If you find any books interesting and suitable for the Data Mining lecture/practical course apart from the suggested, please let us and other students know.