Piotr Bonar

Data Engineering

Labs 3

**Exercise 1:**

A diagram of a mathematical equation

AI-generated content may be incorrect.

**Graph Representation:**

*V*1​ has connections to:

* *V*2​ with weight 3
* *V*3​ with weight 5

*V*3​ has connections to:

* *V*2​ with weight 2

*V*4​ has connections to:

* *V*6​ with weight 6

*V*5​ has connections to:

* *V*4​ with weight 7
* *V*6​ with weight 4.

Yes, it is directed graph, because the weights are different from one to another on both ways

Eg. V4 🡪 V6 has a weight 6, whileV6 🡪 V4 does not have a path at all.

**Exercise 2:**

Store the following matrix using: Coordinate list (COO), Dictionary of

Keys (DOK), compressed space row (CSR), and compressed space column (CSC).

0 0 1 3 0

7 0 0 0 0

0 4 0 0 2

1 0 0 3 5

0 2 7 0 0

1. Coordinate List (COO):

Row Column Value

0 2 1

0 3 3

1 0 7

2 1 4

2 4 2

3 0 1

3 3 3

3 4 5

4 1 2

4 2 7

1. Dictionary of Keys (DOK):

Key Value

(0,2) 1

(0,3) 3

(1,0) 7

(2,1) 4

(2,4) 2

(3,0) 1

(3,3) 3

(3,4) 5

(4,1) 2

(4,2)​ 7

1. Compressed Sparse Row (CSR):

Data: [1,3,7,4,2,1,3,5,2,7]

Row ind.: [2,3,1,2,5,1,3,4,1,2]

Col ptr.: [0,2,3,5,8,10]

​

1. Compressed Sparse Column (CSC):

Data: [7,1,4,2,1,7,3,3,5,7]

Indices: [1,3,2,4,0,4,0,3,2,3]

Indptr: [0,2,4,6,8,10]

**Exercise 3:**

First line:

|  |  |  |
| --- | --- | --- |
|  | Star | Circles |
| Left | 7 | 5 |
| Right | 18 | 20 |

Precision = 0.58, Recall = 0.28, F1-Score = 0.38, Accuracy = 0.54

Second line:

|  |  |  |
| --- | --- | --- |
|  | Stars | Circles |
| Left | 17 | 6 |
| Right | 8 | 19 |

Precision = 0.74, Recall = 0.68, F1-Score = 0.71, Accuracy = 0.72

Third line:

|  |  |  |
| --- | --- | --- |
|  | Stars | Circles |
| Left | 24 | 16 |
| Right | 1 | 9 |

Precision = 0.60, Recall = 0.96, F1-Score = 0.74, Accuracy = 0.66**A graph with different colored dots and numbers

AI-generated content may be incorrect.**

**Exercise 4:**

0-1 Normalized Data:

[[0.03571429 0.57692308 0.67307692 0. 0.04166667 ]

[0.14285714 0.32692308 0. 0.33333333 0. ]

[0.25 0.67307692 0.80769231 0.44444444 1. ]

[0. 0.44230769 0.32692308 0.19444444 0.025 ]

[0.42857143 1. 1. 0.55555556 0.20833333 ]

[0.08928571 0.28846154 0.23076923 0.27777778 0.05833333 ]

[0.33928571 0.71153846 0.86538462 0.47222222 0.16666667 ]

[0.05357143 0.48076923 0.51923077 1. 0.08333333 ]

[0.21428571 0.63461538 0.76923077 0.41666667 0.10833333 ]

[1. 0. 0.28846154 0.16666667 0.01666667 ]]

Mean Normalized Data:

[[-0.21964286 0.06346154 0.125 -0.38611111 -0.12916667]

[-0.1125 -0.18653846 -0.54807692 -0.05277778 -0.17083333 ]

[-0.00535714 0.15961538 0.25961538 0.05833333 0.82916667 ]

[-0.25535714 -0.07115385 -0.22115385 -0.19166667 -0.14583333 ]

[ 0.17321429 0.48653846 0.45192308 0.16944444 0.0375 ]

[-0.16607143 -0.225 -0.31730769 -0.10833333 -0.1125 ]

[ 0.08392857 0.19807692 0.31730769 0.08611111 -0.00416667 ]

[-0.20178571 -0.03269231 -0.02884615 0.61388889 -0.0875 ]

[-0.04107143 0.12115385 0.22115385 0.03055556 -0.0625 ]

[ 0.74464286 -0.51346154 -0.25961538 -0.21944444 -0.15416667 ]]

Z-score Normalized Data:

[[-0.78261442 0.24420337 0.4044764 -1.49635475 -0.45551801 ]

[-0.40085129 -0.71780992 -1.77347347 -0.2045377 -0.6024593 ]

[-0.01908816 0.61420849 0.84006638 0.22606798 2.92413171 ]

[-0.90986879 -0.27380378 -0.7156121 -0.7427948 -0.51429452 ]

[ 0.61718373 1.87222587 1.46233777 0.65667367 0.13224716 ]

[-0.59173285 -0.86581197 -1.0267478 -0.41984054 -0.39674149 ]

[ 0.29904779 0.76221053 1.0267478 0.3337194 -0.01469413 ]

[-0.71898723 -0.12580174 -0.09334071 2.3790964 -0.30857671 ]

[-0.14634253 0.46620644 0.7156121 0.11841656 -0.22041194 ]

[ 2.65325375 -1.97582731 -0.84006638 -0.85044622 -0.54368278 ]]

Outliers (row, column):

[]