Name:	SOLUTION
IIT ID Number:	× × × × × × × × × ×

CS 579: Online Social Network Analysis

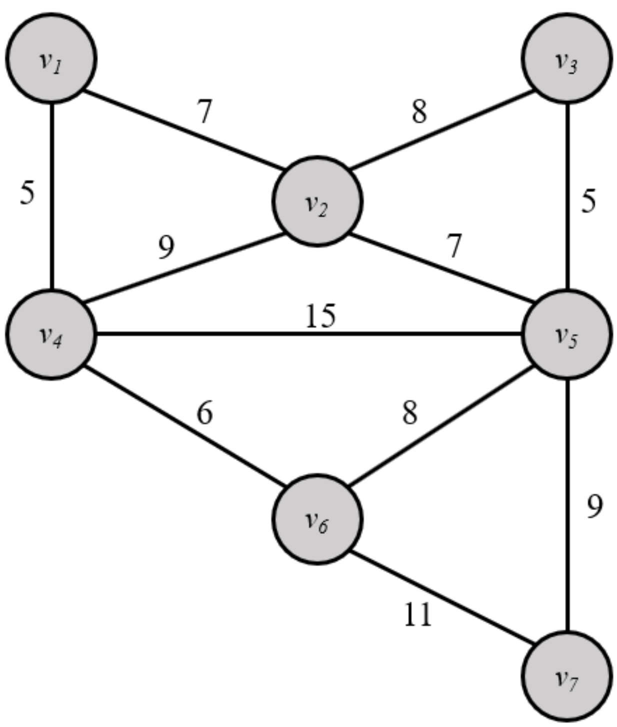
Homework I - Graph Essentials, Data Mining

Prof. Kai Shu
Due at 2022 Feb. 7th, 11:59 PM

This is an ***individual*** homework assignment. Please submit a digital copy of this homework to **Blackboard**. For your solutions, even when not explicitly asked, you are supposed to concisely justify your answers.

1. **[Graph Algorithms]**

   (a) Compute the shortest path between v_1 and other nodes using Dijkstra's algorithm for the following graph.

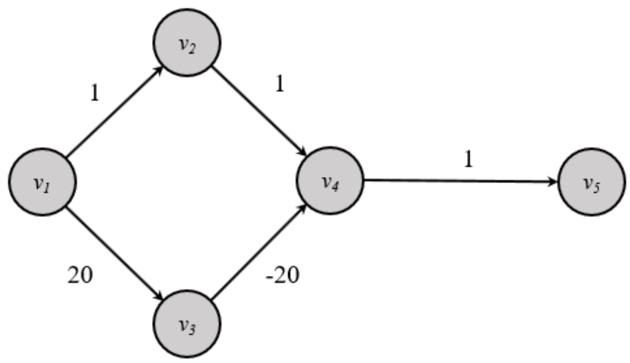


Node	Distance from v_1
v_2	7
v_3	15
v_4	5
v_5	14
v_6	11
v_7	22

Name:	SOLUTION
IIT ID Number:	× × × × × × × × × ××

(b) In the space below, draw a simple example of a directed graph with negative-weight edges for which Dijkstra's algorithm produces incorrect answers.

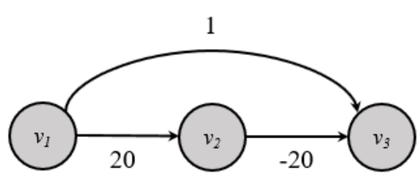
Suppose v_1 is the source node and v_5 is the destination node. The shortest path between v_1 and v_5 is $v_1v_3v_4v_5$, while the Dijkstra algorithm will return $v_1v_2v_4v_5$

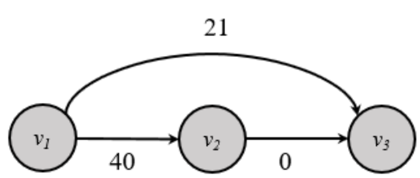


Any other correct counterexamples would be acceptable.

(c) Argue whether "Algorithm 1" below always produces the shortest paths from one source node to others for graphs that have negative weights but do not have negative cycles.

The Dijkstra algorithm will return v_1v_3 as the shortest path in both cases, while the actual shortest path is $v_1v_2v_3$ which this modification did not help to achieve it.





Any other correct counterexamples and arguments would be acceptable.

Algorithm 1: Dijkstra Algorithm for graphs with negative weights.

Input : Adjacency Matrix M, Source node s.

Output: Shortest Path from s to other nodes.

$C \leftarrow$ Find minimum weight in M

for all i and j:

  $M[i,j] \leftarrow M[i,j] - C$

return Dijkstra(M, s) *// use the original Dijkstra algorithm to find the shortest paths*

Name:	SOLUTION
IIT ID Number:	× × × × × × × × × ××

2. **[Network Algorithms]** For a real-world social network, is BFS or DFS more desirable? Why? Provide details.
We rather use BFS if the underlying graph of the social network has high average degree and low diameter. On the contrary, DFS would be more desirable if the graph has either low average degree or high diameter as it could visit the nodes faster.

3. **[Decision Tree and Data Types]** Consider the given dataset below. Answer the following questions:

Instance	Age	Income	Student	Credit Rating	Buy Computer
1	25	High	No	Fair	No
2	20	High	No	Excellent	No
3	32	High	No	Fair	Yes
4	45	Medium	No	Fair	Yes
5	41	Low	Yes	Fair	Yes
6	41	Low	Yes	Excellent	No
7	36	Low	Yes	Excellent	Yes
8	27	Medium	No	Fair	No
9	30	Medium	Yes	Fair	Yes
10	42	Medium	Yes	Fair	Yes
11	29	Medium	Yes	Excellent	Yes
12	31	Medium	No	Excellent	Yes
13	33	High	Yes	Fair	Yes
14	41	Medium	No	Excellent	No

(a) Specify the data types (Nominal, Ordinal, Interval, Ratio) for each of the four attributes (Age, Income, Student, Credit Rating) in the given data.

	Age	**Income**	**Student**	**Credit Rating**
Data Type	Ratio	Ordinal	Nominal	Ordinal

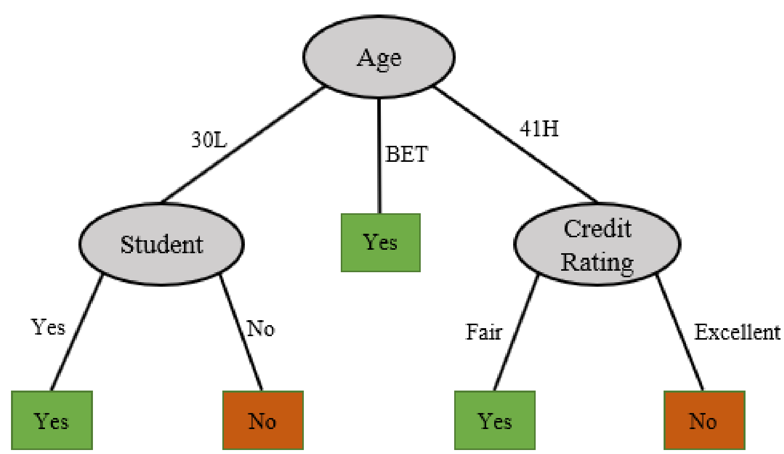
(b) Now assume that we have discretized the real-value "Age" attribute into three categories: 1) 30L: "Age"$\leq$ 30, 2) 41H: "Age"$\geq$ 41, and 3) BET: 31 $\leq$"Age"$\leq$ 40. What is the new data type for the "Age" attribute given this change?

	Age
Data Type	Ordinal

Name:	SOLUTION
IIT ID Number:	× × × × × × × × × × ×

(c) Using the ID3 algorithm that we discussed in the class, generate the decision tree for the given dataset. Assume that "Buy Computer" attribute is the class label and the "Age" attribute is discretized as we discussed in previous question. Note that there could be more than one tree that fits the same data and we only need one! Show all your work for each step in making decision tree and explain how you select decision tree nodes and branches.

Instance	**Age**	**Income**	**Student**	**Credit Rating**	**Buy Computer**
1	30L	High	No	Fair	No
2	30L	High	No	Excellent	No
3	BET	High	No	Fair	Yes
4	41H	Medium	No	Fair	Yes
5	41H	Low	Yes	Fair	Yes
6	41H	Low	Yes	Excellent	No
7	BET	Low	Yes	Excellent	Yes
8	30L	Medium	No	Fair	No
9	30L	Medium	Yes	Fair	Yes
10	41H	Medium	Yes	Fair	Yes
11	30L	Medium	Yes	Excellent	Yes
12	BET	Medium	No	Excellent	Yes
13	BET	High	Yes	Fair	Yes
14	41H	Medium	No	Excellent	No



Name:	SOLUTION
IIT ID Number:	$\times\times\times\times\times\times\times\times\times\times$

4. **[Naive Bayes Classification]** Using the Naive Bayes algorithm and the table given in question 3, what would be the label for the following instance. Assume that "Buy Computer" attribute is the class label and the "Age" attribute is discretized as we discussed in 3.(b).

	Age	**Income**	**Student**	**Credit Rating**	**Buy Computer**
Instance 15	26	Low	Yes	Fair	Yes

We should compare P(Buy Computer="Yes"|Age="30L", Income="Low", Student="Yes", Credit Rating="Fair") with P(Buy Computer="No"|Age="30L", Income="Low", Student="Yes", Credit Rating="Fair").
P(Buy Computer="Yes"|Age="30L", Income="Low", Student="Yes", Credit Rating="Fair") $\propto$ P(Age="30L", Income="Low", Student="Yes", Credit Rating="Fair"|Buy Computer="Yes")$\cdot$P(Buy Computer="Yes")$=\frac{2}{9} \times \frac{2}{9} \times \frac{6}{9} \times \frac{6}{9} \times \frac{9}{14} = \frac{16}{1134} \approx 0.01411$.
P(Buy Computer="No"|Age="30L", Income="Low", Student="Yes", Credit Rating="Fair") $\propto$ P(Age="30L", Income="Low", Student="Yes", Credit Rating="Fair"|Buy Computer="No")$\cdot$P(Buy Computer="No")$=\frac{3}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{14} = \frac{3}{875} \approx 0.00343$