UCLA CS 145 Homework #3

DUE DATE: Wednesday, June 2 2021, 23:59 PM

Note

- Late submissions will generally **NOT** be accepted. Each student has an one-day extension for **ONE of three** homework assignments if the student contacts the instructor and TAs **BEFORE** the submission deadline to arrange the only late submission.
- Discussions on homework assignments are encouraged, but any form of cheating and plagiarism will **NOT** be tolerated. Every student should accomplish the own solutions alone. *Suspicious cases will be reported to The Office of the Dean of Students.*

1 SVM

(20%) Consider classifying the following 20 data points in the 2-d plane with class label y as in Table 1. Suppose by solving the dual form of the quadratic programming of SVM, we can derive the α_i 's for each data point as follows: Among j = 0, 1, $\cdots$, 19 (note that the index starts from 0), 1 = 0.5084, 5 = 0.4625, 17 = 0.9709, and j = 0 for all other j.

ID	x1	x2	y
0	0.52	-1.00	1
1	0.91	0.32	1
2	-1.48	1.23	1
3	0.01	1.44	1
4	-0.46	-0.37	1
5	0.41	2.04	1
6	0.53	0.77	1
7	-1.21	-1.10	1
8	-0.39	0.96	1
9	-0.96	0.08	1
10	2.46	2.59	-1
11	3.05	2.87	-1
12	2.20	3.04	-1
13	1.89	2.64	-1
14	4.51	-0.52	-1
15	3.06	1.30	-1
16	3.16	-0.56	-1
17	2.05	1.54	-1
18	2.34	0.72	-1
19	2.94	0.13	-1

Table 1: Data Points for Problem 1

(a) (5%) Which vectors in the training points are supporting vectors? Explain why.

(b) (10%) Compute the normal vector w and the bias b for the hyperplane.

(c) (5%) With the parameters w and b, we can now use our SVM to do predictions. What is the predicted label of $x_{new} = (2, -0.5)$?

2 Tree Representation

(20%) Given three trees shown in Figure 1, answer the following questions.

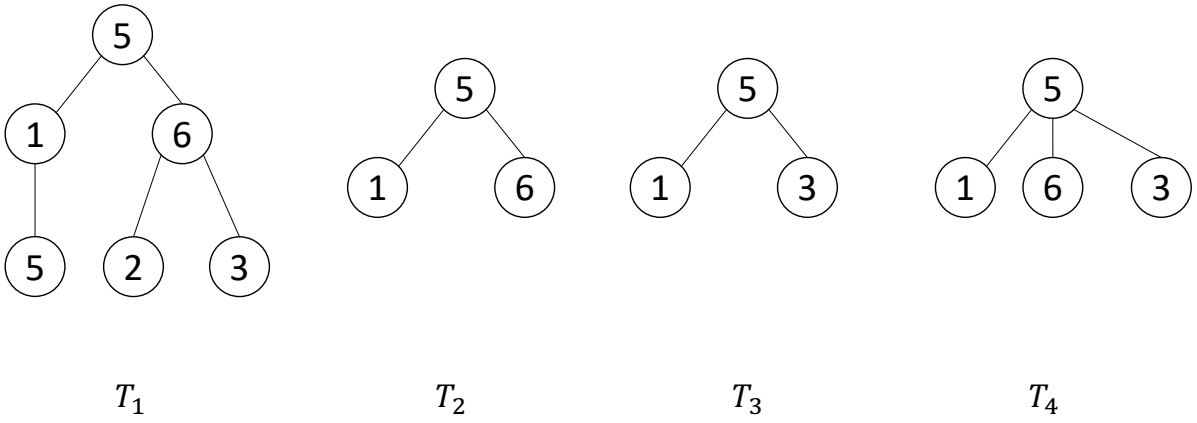

Figure 1: Four Trees for Problem 2.

(a) (10%) Show the string representation of each tree with a DFS traversal.

(b) (5%) Judge whether T_2, T_3, T_4 are embedded subtrees of T_1?

(c) (5%) Judge whether T_2, T_3, T_4 are induced subtrees of T_1?

3 Tree Mining

(20%) Given three trees shown in Figure 2 and `min_support` $= 2$, apply the *Apriori Style TreeMiner* algorithm to find all frequent subtrees (Please write down the string representation for each frequent subtree).

4 Graph Mining

(40%) Given three graphs shown in Figure 3, answer the following questions given `min_support` $= 2$. (Assume $0 < a < b < c < \cdots < x < y$.)

(a) (10%) Derive the **minimum** DFS code representation for each graph.

(b) (10%) Sort the **minimum** DFS code representation of three graphs in a DFS lexicographic order.

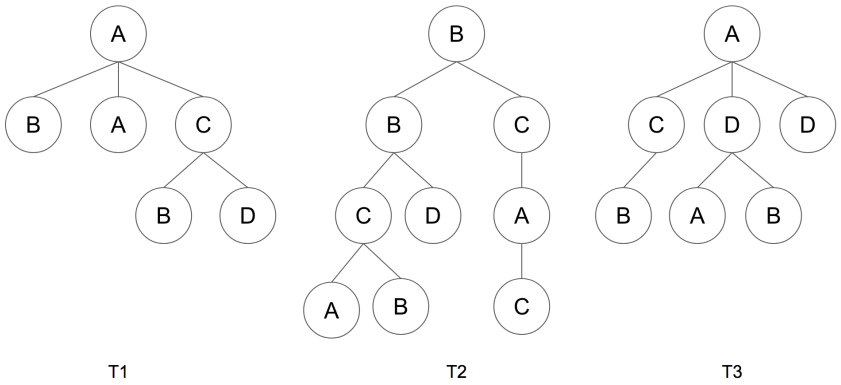

Figure 2: Three Trees for Problem 3.

(c) (15%) Apply the *gSpan* algorithm to find all frequent subgraphs. Suppose we do **not** count single-node graphs as frequent subgraphs. How many frequent subgraphs are there in total?

(d) (5%) Derive the Canonical Adjacency Matrix (CAM) of G_1.

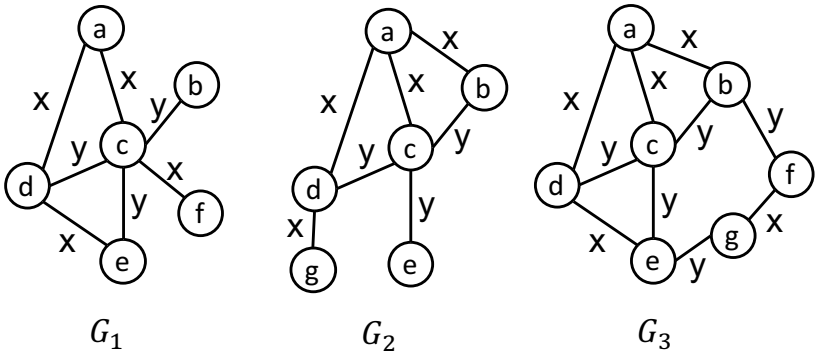

Figure 3: Three graphs for Problem 4.