

Homework 5

(due - 23:59 June 18, 2022)

Please submit your homework as a PDF file.

Questions:

[1] Study Supervised Learning (SL) and Unsupervised Learning (UL) and answer the following questions (NB: Your answer of the three questions altogether should not exceed one page.) [20 points]

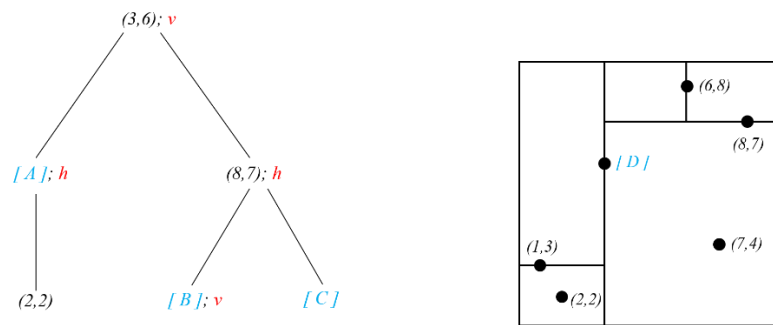
- a. Write down the difference between SL and UL. [10 points]
 - A. Supervised learning needs the desired solutions which are called labels for the training. By given input features, the model predicts the output and compares to the label for updating the model. Therefore, supervised learning is widely used in classification problem, which labels are categorical(nominal), and regression problem, which labels are ordinal number. Not like the supervised learning, unsupervised learning does not need labeling on the dataset. It uses the internal features for the training. Therefore, unsupervised learning is widely used in clustering which combine the data with similar characteristics, and anomaly detection.
- b. List two examples of SL. [5 points]
 - A. Linear regression, Neural networks
- c. List two examples of UL. [5 points]
 - A. Clustering, Anomaly detection

[2] For each question below, choose your answer. [20 points]

- a. Supervised Learning (SL) is learning with labeled data, and examples of SL are learning representations and clustering. [True / False]
- b. If we apply the k-means clustering algorithm with 15 instances and 3 clusters, each cluster will always have exactly 5 assigned instances as a result of clustering. [True / False]
- c. In k-means clustering algorithm, If only three instances on the xy-plane A(0,0), B(0,1), C(1,0) are assigned to a cluster X, the recomputed cluster center of X is (1,1). [True / False]
- d. Decision tree stop develop the branch of the tree if all instances at a node have the same class. [True / False]
- e. Deep Q-Networks algorithm can learn stochastic policy. [True / False]

[3]

(a)



The tree (left) and map on a 2d plane (right) in the figure above represent a same kD-tree. Find correct coordinates for $[A] \sim [D]$.

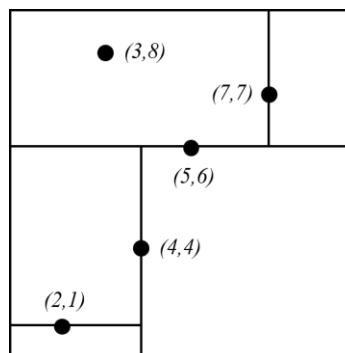
[A] = (1,3)

[B] = (6,8)

[C] = (7,4)

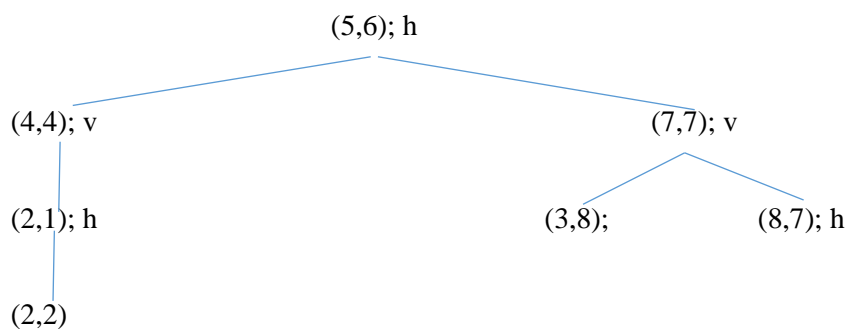
[D] = (3,6)

(b)



Draw a tree that has same information in the map above. For your tree, each node must have coordinate pair and corresponding **split option** among *horizontal* (*h*) / *vertical* (*v*) / *none* (*blank*). For more detailed explanations for split options, see the tree in (a). Then you can check [30 points]

- Node (8,7)'s option is horizontal (*h*) and it has horizontal line on $y=7$.
- Node (2,2)'s option is none (*blank*) and there is no line that go through (2,2).



[4] The following table shows a sample with features x_1 , x_2 , and Actual labels.

The weights and biases for the two models (decision boundary & logistic regression) trained with this data are as follows:

Table 1. Sample Datasets

	x_1	x_2	Actual Label
1	2	9	True
2	2	0	False
3	3	8	True
4	1	0	False
5	7	3	False

Table 2. The Weights and biases for each model.

Model	W_1	W_2	bias	Threshold
Decision Boundary	0.3	0.7	-2.0	Over 0 is True
Logistic Regression	0.7	0.4	-5.0	Over 0.5 is True

Calculate the Accuracy, precision, recall, and f1-score of each model. [30 points]

a) $Y = W_1X_1 + W_2X_2 + B$

A. Prediction: TFTFT, (4.9, -1.4, 4.5, -1.7, 2.2)

B. TP – 2, TN – 2, FP – 1, FN – 0

b) $Y = \exp^{(W_1X_1 + W_2X_2 + B)} / (1 + \exp^{(W_1X_1 + W_2X_2 + B)})$

A. Prediction: FFTFT, (0.5, 0.026597, 0.574443, 0.013387, 0.75026)

B. TP – 1, TN – 2, FP – 1, FN – 1

Answer sheet: give your answer to 2 decimal places

Model	Accuracy	Precision	Recall	F1 Score
Decision Boundary	0.8	2/3	1	0.8
Logistic Regression	0.6	0.5	0.5	0.5

(Posted on 2022/06/07)