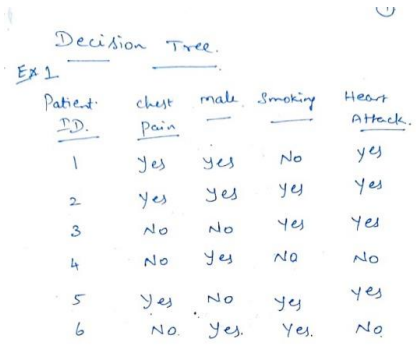
Worksheet for Module 3 & 4

Question: 21 **Decision tree**



Decision Tree.

Ex 1.

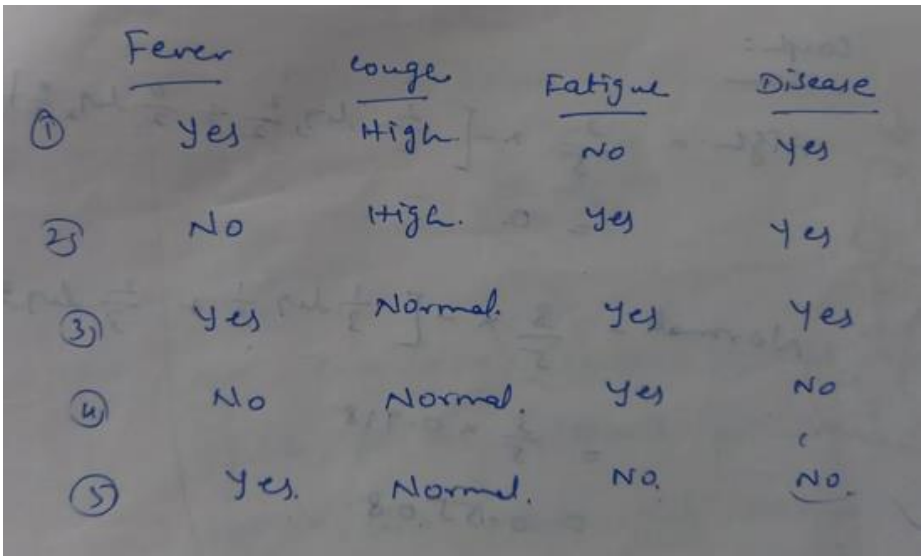
Patient ID.	chest Pain	male	Smoking	Heart Attack.
1	Yes	Yes	No	Yes
2	Yes	Yes	Yes	Yes
3	No	No	Yes	Yes
4	No	Yes	No	No
5	Yes	No	Yes	Yes
6	No	Yes	Yes	No

Construct a decision tree for the above data and classify the input features

"Chest pain= No Male=yes Smoking =yes Heart attack=?"

Question:22 kNN

Apply kNN algorithm on the following dataset and find the predict the target class for the following feature "fever=No cough=normal fatigue=yes" with k=2



	Fever	Cough	Fatigue	Disease
1	Yes	High	No	Yes
2	No	High	Yes	Yes
3	Yes	Normal	Yes	Yes
4	No	Normal	Yes	No
5	Yes	Normal	No	No

Question:23 SVM

Apply linear SVM algorithm for the following data and find the support vectors and linearly separable line.

x	y	T
0	2	+ve
0	-1	+ve
2	0	+ve
4	0	-ve
5	3	-ve
6	-2	-ve

Question: 24 (kNN with bagging)

Use bagging with kNN for the following dataset for three bootstrap sample with k=2. For a new test sample X1=4.0,X2=4.5,X3=2.2 find the target class value?

Bootstrap sample1: (1,2,5,6,7)

Bootstrap sample 2: (3,4,5,8,10)

Bootstrap sample 3: (1,3,6,7,9)

Sample	Feature X1	Feature X2	Feature X3	Target Y
1	2.0	3.1	1.5	0
2	3.2	4.3	2.1	1
3	1.5	2.8	1.2	0
4	4.1	3.9	2.5	1
5	5.3	6.7	3.2	1
6	3.8	5.1	2.8	1
7	2.7	3.5	1.8	0
8	4.5	4.9	2.7	1
9	1.9	3.0	1.4	0
10	5.1	6.2	3.0	1

Question 25: Random forest

A classification problem where we predict whether a person will buy a product based on three attributes:
Age (Young, Middle, Old)
Salary (Low, Medium, High)
Online Ads Clicked (Few, Moderate, Many)
Target Variable (Buy Product?): Yes or No

ID	Age	Salary	Clicks	Buy?
1	Young	Low	Few	No
2	Young	High	Many	Yes
3	Middle	Medium	Moderate	Yes
4	Old	Low	Few	No
5	Old	Medium	Many	Yes
6	Middle	High	Few	No
7	Young	Medium	Moderate	Yes
8	Middle	Low	Few	No
9	Old	High	Many	Yes
10	Young	Low	Many	Yes

Build 3 Decision trees using the three bootstrap samples and classify the features using random forest algorithm

Age=Old, Salary=High, Clicks=Few, Buy=?

Tree 1:
Sample: {1, 3, 4, 5, 7, 9}
Tree 2:
Sample: {2, 4, 6, 8, 9, 10}
Tree 3:
Sample: {1, 3, 5, 6, 7, 10}

Question 26: AdaBoost

Assume that third and fifth samples are misclassified, so to strengthen this weak samples apply Adaboost algorithm and generate the new dataset for the next iteration. Assume the random weights are 0.1, 0.5, 0.7, 0.8, 0.9, 0.7, 0.6, 0.85. The samples in new dataset can be duplicated.

F1	F2	F3	O/P
12	3	23	YES
23	5	45	YES
34	3	43	NO
21	4	65	YES
45	5	34	NO
12	2	23	NO
34	5	43	YES
16	6	45	YES

Question 27: Multiclass using One Vs ALL and One Vs One

Apply both the multiclass approach on the given dataset and find the class label for the test sample “color=red shape=oval taste =bitter”

Color	Shape	Taste	Fruit
Red	Round	Sweet	Apple
Yellow	Round	Bitter	Lemon
Yellow	Oval	Sweet	Mango
Red	Oval	sweet	Apple
Yellow	Round	sweet	lemon

Question 28: Error correcting output codes

For the Question No. 26 apply the binary codes to decide whether the output codes need to be flipped or not