

Some practice questions:

- Given the following table, create a random forest of 4 decision trees, with two features x1,x3 and x2,x4 and 4 rows for each DT, taken in perfect shuffle. What is the prediction for the new entry?

Sl. No.	X1	X2	X3	X4	Decision
1	1	A	Z	5	Y
2	2	B	Y	4	Y
3	5	A	Z	8	Y
4	2	C	Q	2	N
5	4	A	R	3	Y
6	3	F	R	5	N
7	5	B	R	4	Y
8	3	A	Q	3	Y
9	2	B	Z	3	Y
10	1	B	Y	8	Y
11	5	A	Z	8	Y
12	2	C	Z	2	N
13	5	A	Y	3	Y
14	4	F	R	8	N
15	5	B	Y	4	Y
16	4	A	Q	3	Y
New	4	B	Y	3	?

- PCA: What does the Eigen value denote? What does the values of an Eigen vector denote? Illustrate how we can obtain the transformed dataset using a single principal component. If we choose a threshold of 90% for selecting principal components, what formula will you apply for selecting them? What would be the loss of information?
- Given height and gender, find probability of selecting a new female of 170 cm height. Use NB.
Height distribution of selected candidates: average 160 cm, standard dev: 15 cm
Height distribution of rejected candidates: average 150 cm, standard dev: 10 cm
7 females and 9 males were selected. 5 females and 20 males were rejected.
- In ADABOOST, 3 iterations were performed. In the first iteration, 3 out of 10 data points were erroneous (say d1, d2, d3). In the second round, only 2 of these data points were still erroneous (say d1, d3). In the third round only 1 of them d1 was erroneous. Show the weight distribution of the data points after each round. Show the final prediction of a data point that predicted -1, -1, 1 in the three rounds.
- Given the BW pixel map and kernel below, Calculate the output of first stage of CNN after applying valid padding vertically and horizontally with stride 1 in both dimensions, convolution operation, ReLU activation, then maxpooling.

1	1	1	0	1	0	0	1	0
1	1	0	1	0	1	1	1	0

1	1	1	0	0	0
1	0	1	0	1	0

6.