Exercises on K-nearest Neighbor and Decision Trees

Consider the following dataset:

ID	X1	X2	Y	ID	X1	X2	Y
1	-1	-1	2	7	1	-1	2
2	-1	2	1	8	1	0	2
3	-1	4	1	9	2	0	3
4	0	-1	1	10	2	3	3
5	0	0	1	11	-1	0	?
6	0	3	2	12	1	2	?

Using K-Nearest Neighbor classifier with Manhattan Distance, what are the predicted labels of points 11 and 12? Starting with K=1, increase K till there are no ties. If there are multiple neighbours at equal distance, consider all of them. Assume that all neighbors have equal weightage.

Consider the following dataset:

ID	X1	X2	Y	ID	X1	X2	Y
1	2	3	A	7	0	-1	C
2	2	0	A	8	-1	4	C
3	1	0	B	9	-1	2	C
4	1	-1	B	10	-1	-1	B
5	0	3	B	11	-1	0	?
6	0	0	C	12	1	2	?

Using K-Nearest Neighbor classifier with Manhattan Distance, what are the predicted labels of points 11 and 12? Starting with K=1, increase K till there are no ties. If there are multiple neighbours at equal distance, consider all of them. Set the weightage of each neighbor as reciprocal of distance from query point

In case of the following dataset, you want to build a decision stump. Which splitting criteria will you choose?

ID	X1	X2	Y	ID	X1	X2	Y
1	1	5	A	6	8	2.5	A
2	6	5	B	7	5	2	A
3	2	4	A	8	3	1.5	B
4	6	4	B	9	1	1	B
5	1	2.5	B	10	6	0	A

a) X1>4 b) X1>5.5 c) X2>3 d) X2>0.5

Consider the following dataset:

ID	1	2	3	4	5	6	7	8
X1	1	1	1	1	2	2	2	2
X2	1	1	2	2	1	1	2	2
X3	1	2	1	2	1	2	1	2
Y	3	10	4	12	2	9	2	?

We want to build a regression stump, using variance reduction as the splitting criteria (original variance – split1fraction*split1variance – split2fraction*split2variance)

On which attribute will you split?

a) X1 b) X2 c) X3 d) all equivalent