

HOMEWORK #2

1)

(a) Which attribute is deemed to be the most discriminative one for class prediction?

According to the result obtains after the simulation, the most discriminative attribute is

Node-caps because It's the root of the decision tree as we can see in the following screenshot.



(b) What is the height of the Decision Tree generated?

-The height of the decision tree is 6.

c) Find a pure partition in the Decision Tree and report a screenshot that shows the example identified

-Unfortunately i could not find a pure partition in the decision tree

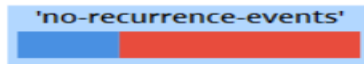
2) Impact of minimal gain and maximal depth parameters on the characteristics of the decision tree model

a) maximal depth=10 and minimal gain=0.01

varying only the maximal depth(from 10 to 5) we can notice how the numbers of nodes in the decision tree decreases significantly as shown in the following picture.



b) maximal depth =10 and minimal gain = 0.1



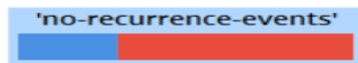
c) maximal depth = 5 and minimal gain = 0.01



d) minimal gain = 0.01 and maximal depth= 1000



e) minimal gain = 100 and maximal depth=10



Conclusion:

-when we decrease the value of the minimal gain the numbers of nodes in decision decreases significantly which means the classification is not so precise.

-when we increase the value of the maximal depth, the number of nodes in the decision tree also increases significantly which means that the classification is more precise.

3)

a) maximal depth=10 and minimal gain=0.01

accuracy: 67.44%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	5	7	41.67%
pred. 'no-recurrence-events'	21	53	71.62%
class recall	19.23%	88.33%	

b) maximal depth =10 and minimal gain = 0.1

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	0	0	0.00%
pred. 'no-recurrence-events'	26	60	69.77%
class recall	0.00%	100.00%	

c) maximal depth = 5 and minimal gain = 0.01

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	5	5	50.00%
pred. 'no-recurrence-events'	21	55	72.37%
class recall	19.23%	91.67%	

d) minimal gain = 0.01 and maximal depth= 1000

accuracy: 67.44%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	5	7	41.67%
pred. 'no-recurrence-events'	21	53	71.62%
class recall	19.23%	88.33%	

e) minimal gain = 100 and maximal depth=10

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	0	0	0.00%
pred. 'no-recurrence-events'	26	60	69.77%
class recall	0.00%	100.00%	

Conclusion about the accuracy:

The accuracy increases when we increase the value of the minimal gain until a certain value where it has no effect on the accuracy which remains constant.

The accuracy increases when we decrease the value of the maximal depth but over a certain value of the maximal depth, the accuracy remains constant.

4)

a) K= 5

☒ Table View ☐ Plot View

accuracy: 72.09%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	5	3	62.50%
pred. 'no-recurrence-events'	21	57	73.08%
class recall	19.23%	95.00%	

b) K= 20

☒ Table View ☐ Plot View

accuracy: 70.93%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	2	1	66.67%
pred. 'no-recurrence-events'	24	59	71.08%
class recall	7.69%	98.33%	

c) K= 80

☒ Table View ☐ Plot View

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	0	0	0.00%
pred. 'no-recurrence-events'	26	60	69.77%
class recall	0.00%	100.00%	

d) K= 200

☒ Table View ☐ Plot View

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	0	0	0.00%
pred. 'no-recurrence-events'	26	60	69.77%
class recall	0.00%	100.00%	

e)K=1000

☒ Table View ☐ Plot View

accuracy: 69.77%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	0	0	0.00%
pred. 'no-recurrence-events'	26	60	69.77%
class recall	0.00%	100.00%	

Conclusion about the correlation between the value of K and the accuracy:

- when K increases, the efficiency decreases

****Result obtain using a naïve baise:pic**

☒ Table View ☐ Plot View

accuracy: 73.26%

	true 'recurrence-events'	true 'no-recurrence-events'	class precision
pred. 'recurrence-events'	10	7	58.82%
pred. 'no-recurrence-events'	16	53	76.81%
class recall	38.46%	88.33%	

Since the efficiency obtained using the Naïve Bayes is greater than one obtained using the K-NN, we can conclude that the naïve classifier is actually better than the K-NN on this data set.

5) correlation matrix achieved

Attribut...	age	menopa...	tumor-s...	inv-nodes	node-ca...	deg-mal...	breast	breast-...	irradiat
age	1	0.241	-0.045	-0.001	0.052	-0.043	0.067	-0.024	-0.011
menopa...	0.241	1	0.019	-0.011	0.130	-0.161	0.077	-0.096	-0.075
tumor-size	-0.045	0.019	1	-0.131	0.058	0.133	-0.022	-0.056	-0.022
inv-nodes	-0.001	-0.011	-0.131	1	-0.465	-0.213	0.040	0.063	0.399
node-caps	0.052	0.130	0.058	-0.465	1	0.098	0.024	-0.036	-0.197
deg-malig	-0.043	-0.161	0.133	-0.213	0.098	1	-0.073	0.018	-0.074
breast	0.067	0.077	-0.022	0.040	0.024	-0.073	1	0.175	-0.019
breast-q...	-0.024	-0.096	-0.056	0.063	-0.036	0.018	0.175	1	-0.005
irradiat	-0.011	-0.075	-0.022	0.399	-0.197	-0.074	-0.019	-0.005	1

a) the naïve independence does not actually hold for the breast data set

b) the pair of most correlated attribute is (inv-nodes, nodes-caps) because the highest absolute correlation coefficient.