AI Assignment-3 Report

Theory

Q1.

(i)

Q1 Defining Random variables:-

T: Person has Travelled

C: Person has caught corona

O: Disease other than corona

D: Person has died.

M: Mild case of disease

S: severe case of disease.

(i) (a) $P(T \wedge (C \vee O)) = 0.825$

(b) $P(M \wedge C | T) = 0.150$

(c) $P(C \wedge S | T) = 0.220$

(d) $P(D \wedge O | T) = 0.240$

(e) $P(\neg T \wedge C \wedge S) = 0.025$

(f) $P(S | \neg T) = 0.457$

(g) $P(D \wedge C) = 0.059$

(h) $P(M \vee S) = 0.700$

(i) $P(T | S) = 0.800$

(j) $P(C | T) = P(C | \neg T) = 0.500$, $P(C) = 0.500$

(ii)

1) As probabilities are greater than or equal to 0 and less than or equal to 1 so they create a valid probability distribution.

They follow below axioms:-

→ Non-Negativity $P(A) \geq 0 \ \forall A$

→ Normalization sum of all probabilities is 1.

→ Additivity

iii)

T	C	O	D	M	S	Probabilities
1	1	0	0	1	0	$P(C \wedge M \mid T) = 0.15$
1	1	0	0	0	1	$P(C \wedge S \mid T) = 0.22$
1	0	1	1	0	0	$P(D \wedge O \mid T) = 0.27$
0	1	0	0	0	1	$P(C \wedge S \wedge \neg T) = 0.025$
0	0	0	0	0	1	$P(S \mid \neg T) = 0.457$
0	1	0	1	0	0	$P(D \wedge C) = 0.059$
1	0	0	0	0	1	$P(T \mid S) = 0.800$

iv) Conditional independence for some variables:
$P(C \wedge M \mid T) = 0.15$, $P(C \wedge S \mid T) = 0.22$
If caught corona & severity of disease are conditionally independent given Travelled.

Q2. (a) If we do not switch the choice the probability of winning the key would be 1/3 so in this case it does not matter what door the person reveals, P (winning key | Not switching) = 1/3

If we choose to switch our choice:

If we choose door 1 (loose life) then the other person will reveal door 2 for sure with probability 1 then we will switch to the next door which will surely contain the key.

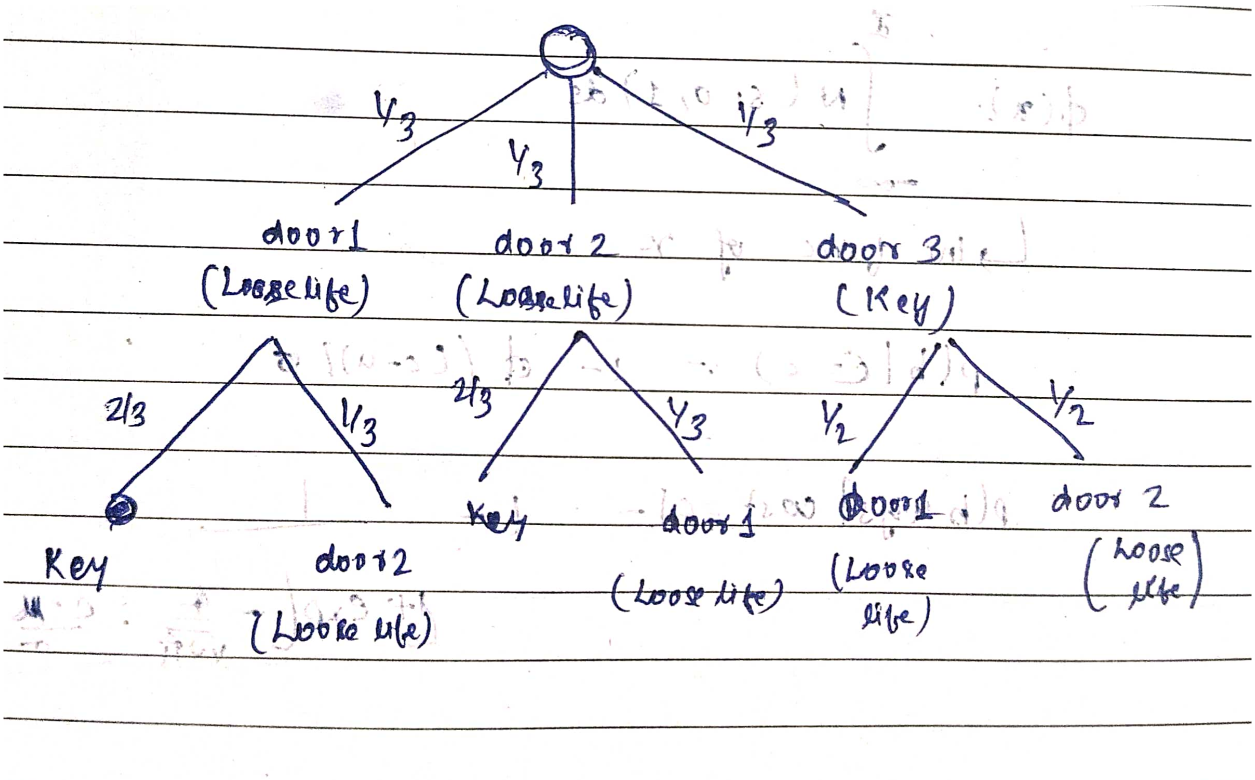
If we choose door 2 (loose life) then the other person will reveal door 2 for sure with probability 1 then we will switch to the next door which will surely contain the key.

If we choose door 3 (Key) then switching makes it impossible to find keys in other doors.

Hence in case of switching P(winning key | Switching) = 1*(1/3) + 1*(1/3) + 0*(1/3) = 2/3

As the probability of finding the key in switching is larger we would choose to switch our choice.

(b)



In case of not switching probability of winning the key remains 1/3

if we choose to switch Probability of the winning key would be calculated from the above diagram as:

P(winning key | Switching) = (1/3)*(1/3) + (1/3)*(1/3) + 0*(1/3) = 2/9

Since P(winning key | Switching) < P(winning key | Not Switching) in this case, we would not switch our choice.

(c)

P(winning Key | Person mistakenly revealed the door that shows the life host)

P(winning key | Reveal door 1 lost life $\vee$ Reveal door 2 lost life)

$$\frac{P(\text{winning key} \wedge (\text{Reveal door 1 lost life} \vee \text{Reveal door 2 lost life}))}{P(\text{Reveal door 1 lost life} \vee \text{Reveal door 2 lost life})}$$

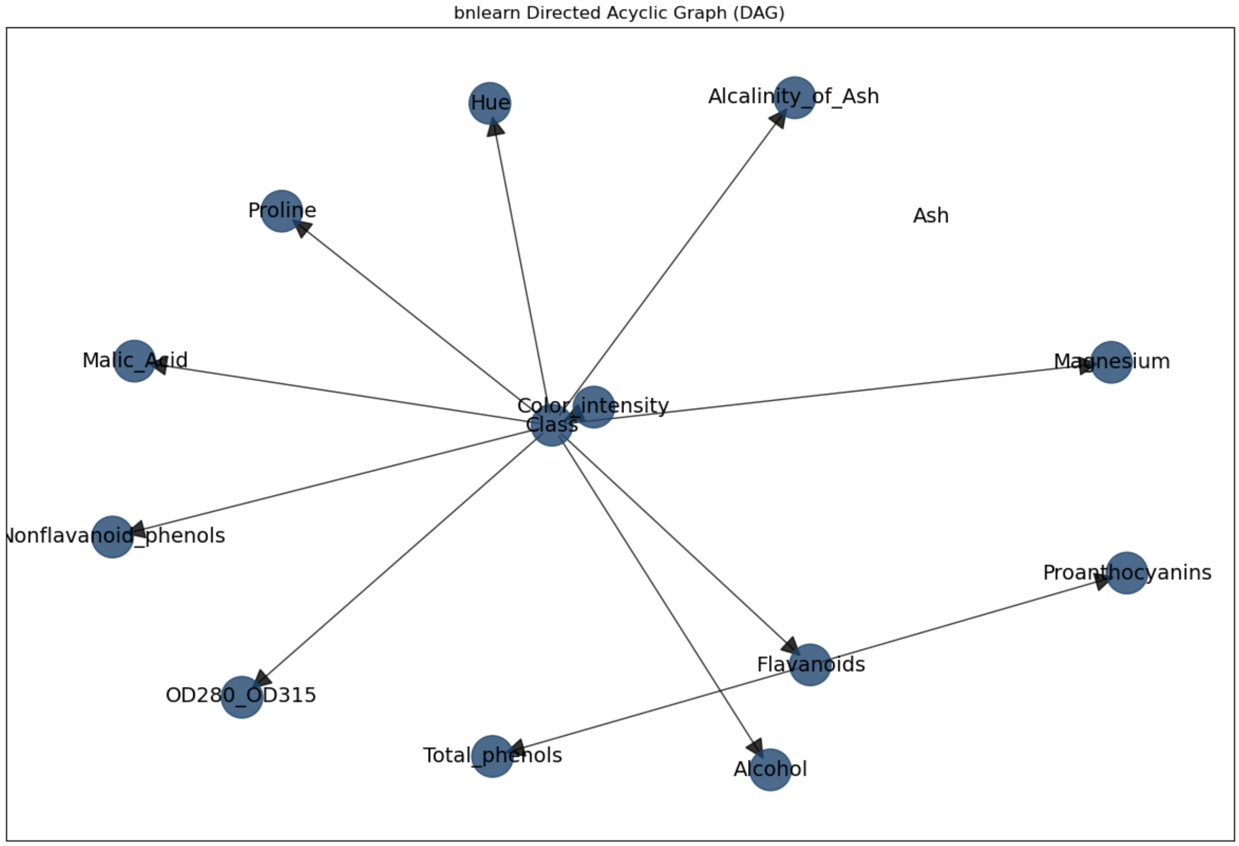
$$\frac{P(\text{Reveal door 2} \mid \text{choose door 1})\,P(\text{choose door 1}) + P(\text{Reveal door 1} \mid \text{choose door 2}) \cdot P(\text{choose door 2})}{P(\text{reveal d2} \mid \text{choose d1})\,P(\text{choose d1}) + P(\text{reveal d1} \mid \text{choose d2}) \cdot P(\text{choose d2}) + P(\text{reveal d2} \mid \text{choose key}) \cdot P(\text{choose key}) + P(\text{reveal d1} \mid \text{choose key}) \cdot P(\text{choose key})}$$

$$\frac{(1/3)(1/3) + (1/3)(1/3)}{(1/3)(1/3) + (1/3)(1/3) + (1/2)(1/3) + (1/2)(1/3)} = \frac{2/9}{5/9} = \frac{2}{5} = 0.4$$

Computational

(a) Loaded the dataset, assigned the column names, and performed the discretization using the cut function from the pandas library, Constructed bayesian network A using structure learning and parameter learning from bnlearn package

Network Visualization:



A few examples of parent and child from the above DAG:

(**Parent—>Child**)

Flavanoids—>Total_Phenols

Flavanoids—>Proanthocyanins

Class—>Malic_Acid

Class—>Hue

All pairs are visible in the above DAG.

(b) Use bn.independence_test and Computed edge strength using chi-square independence test and remove (prune) the not-significant edges (edges with minimum chi-square)

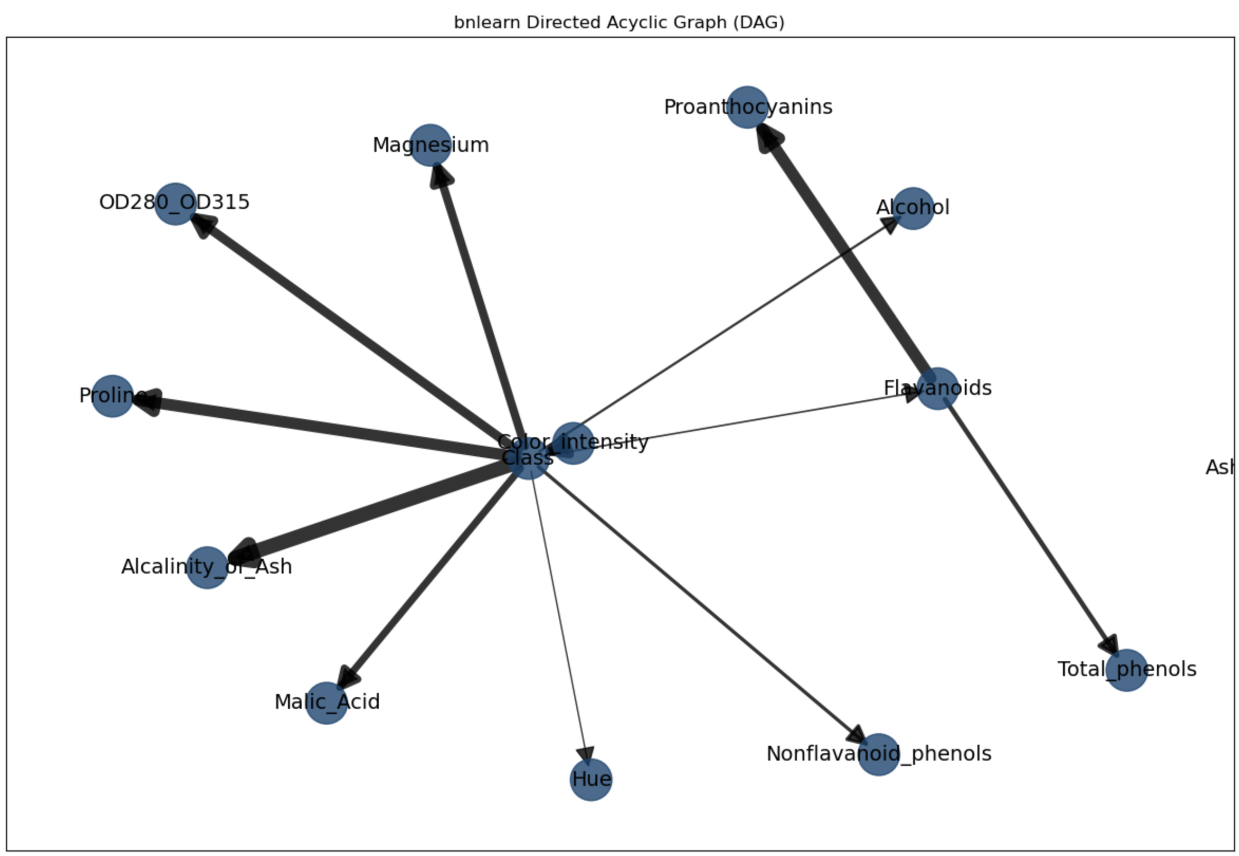
The Chi-square test of independence is a statistical hypothesis test used to determine whether two categorical or nominal variables are likely to be related.

	source	target	stat_test	p_value	chi_square	dof
0	Class	Flavanoids	True	2.76136e-39	194.535	6
1	Class	Proline	True	1.68901e-32	162.572	6
2	Class	OD280_OD315	True	2.5545e-27	138.074	6
3	Class	Hue	True	2.10061e-25	128.987	6
4	Class	Alcohol	True	6.96557e-23	116.995	6
5	Class	Alcalinity_of_Ash	True	1.72727e-12	66.945	6
6	Class	Malic_Acid	True	7.33944e-15	78.4857	6
7	Class	Magnesium	False	4.99726e-11	59.7764	6
8	Class	Nonflavanoid_phenols	False	2.10021e-10	56.7009	6
9	Flavanoids	Total_phenols	True	4.46383e-33	175.472	9
10	Flavanoids	Proanthocyanins	True	1.36322e-17	100.308	9
11	Color_intensity	Class	True	1.08448e-26	135.097	6

As clear from the above we got **chi-square** minimum for Magnesium and Nonflavanoid_phenols hence we will remove those edges from the network to prune the network A.

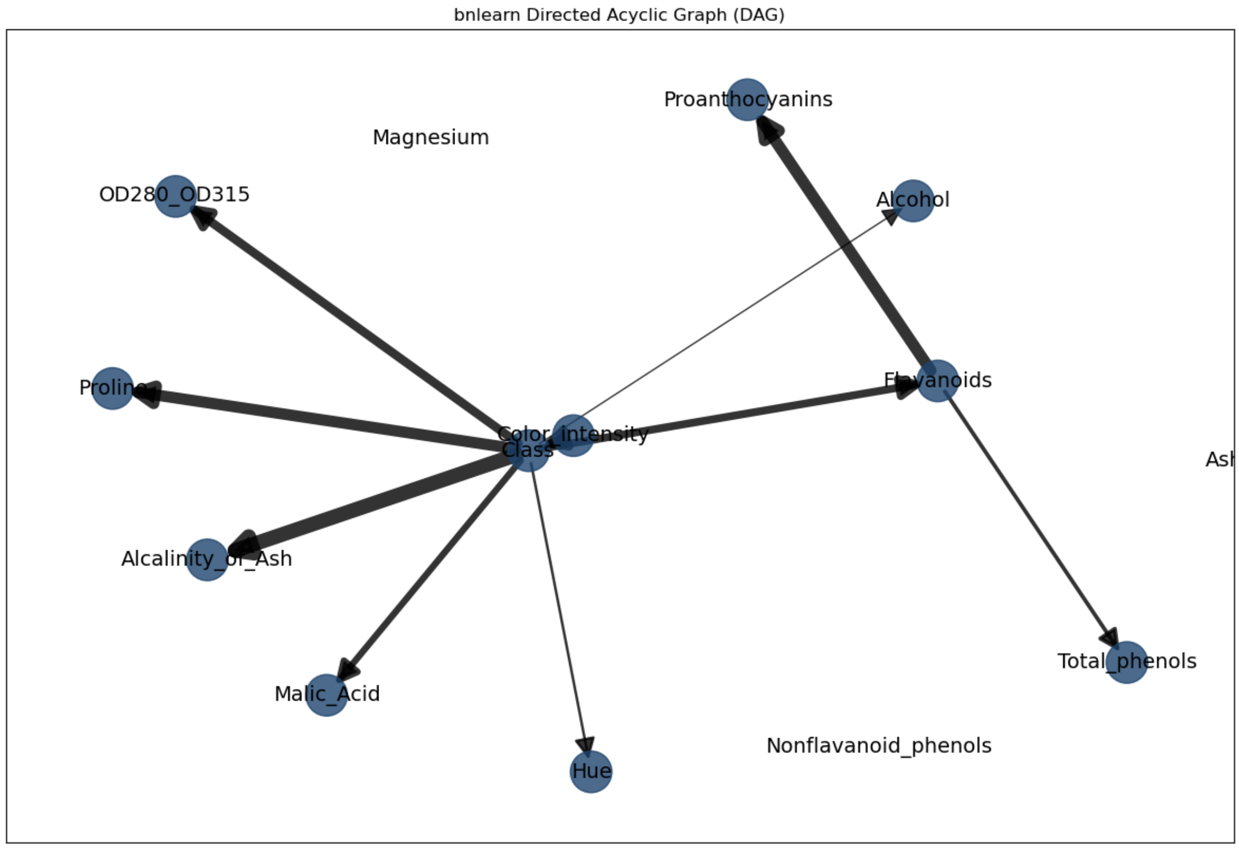
Network before pruning:

Accuracy: 0.9166666666666666



Network after pruning:

Accuracy: 0.9444444444444444

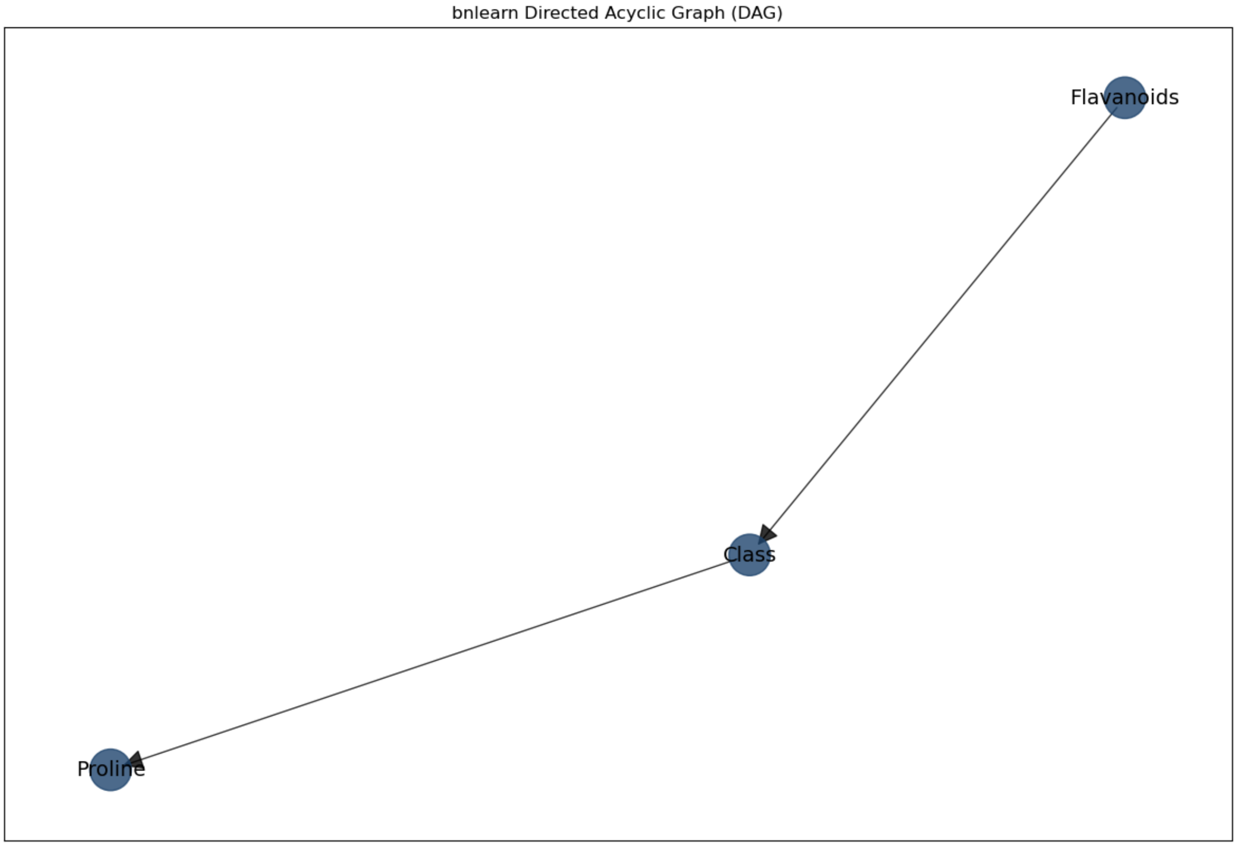


Also, pruned network B has better performance (larger accuracy) than network A.

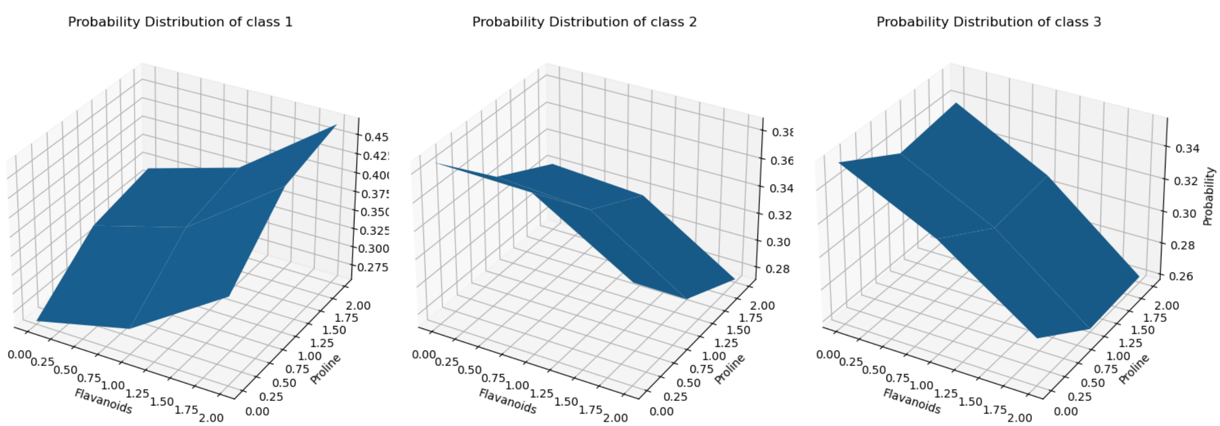
(c) For methods other than pruning, tried different values of parameters to build the model and got the best model with method type as **hill climbing** for structure learning and **maximum-likelihood** for parameter learning with a maximum accuracy of 0.96292929.

(d) Used SelectKBest and f_classif for feature selection, got Featute F1 and F2 as ‘Flavanaoids’ and ‘Proline’ respectively, that are most significant for the class prediction.

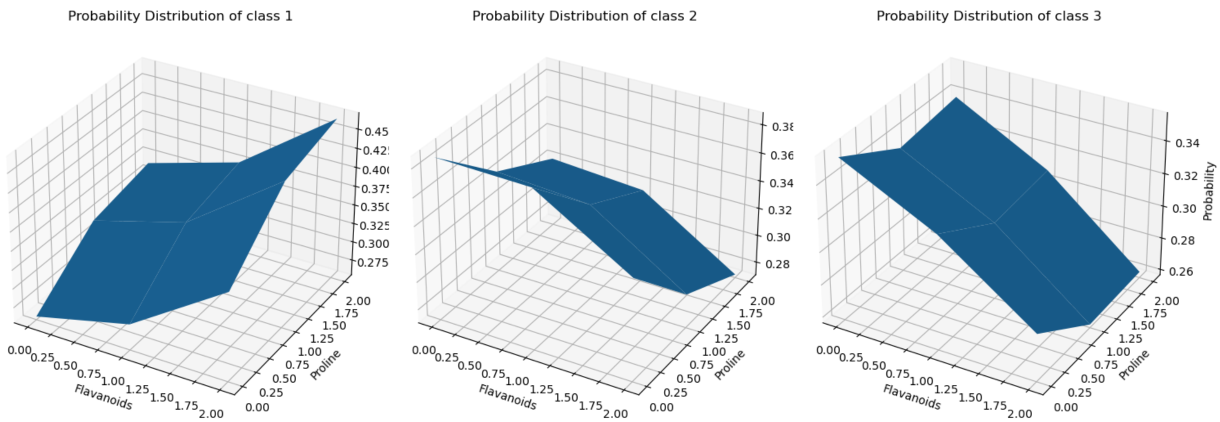
Network C:



The probability distribution for these features with three classes with model C are:



The probability distribution for these features with three classes for model A in the first part are:



As we can observe there is no change in the probability distribution which is simply because the features we have chosen are themselves most significant to the predictions and hence other features won't affect the probability distribution.

Performance of three Bayesian networks:

Bayesian Network	Accuracy
A (original)	0.91666666666
B (Pruned)	0.94444444444
C (feature selection)	0.94444444444

Both Model B and Model C outperform Model A, indicating that pruning and feature selection have positively impacted predictive accuracy.

Posterior probabilities:

```
Posterior probabilities for Class given evidence: +----------+-------------+
| Class    |   phi(Class) |
+==========+==============+
| Class(1) |       0.2602 |
+----------+--------------+
| Class(2) |       0.3875 |
+----------+--------------+
| Class(3) |       0.3523 |
+----------+--------------+
Posterior probabilities for Color_intensity given evidence: +-------------------+----------------------+
| Color_intensity    |   phi(Color_intensity) |
+===================+========================+
| Color_intensity(0) |                0.3321 |
+-------------------+------------------------+
| Color_intensity(1) |                0.3329 |
+-------------------+------------------------+
| Color_intensity(2) |                0.3350 |
+-------------------+------------------------+
Posterior probabilities for Proline given evidence: +------------+---------------+
| Proline    |   phi(Proline) |
+============+================+
| Proline(0) |         0.3281 |
+------------+----------------+
| Proline(1) |         0.3321 |
+------------+----------------+
| Proline(2) |         0.3398 |
+------------+----------------+
Posterior probabilities for Magnesium given evidence: +--------------+-----------------+
| Magnesium    |   phi(Magnesium) |
+==============+==================+
| Magnesium(0) |           0.3364 |
+--------------+------------------+
| Magnesium(1) |           0.3326 |
+--------------+------------------+
| Magnesium(2) |           0.3310 |
```

Yes, these probabilities match my intuition as the highest probability is for the class for which the highest number of data points are present in the data.

The posterior probability is highest for the class for which its count is highest.