Waris Damkham Sec 1 6388014

ITCS451: Artificial Intelligence Assignment
III: Are you ready for the Final?

Part I: CSP

Consider the wolf, goat, and cabbage problem (WGC):

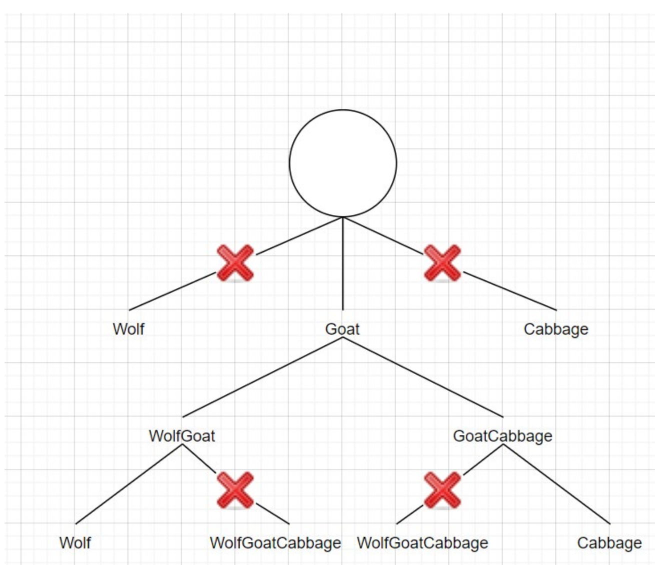
"A farmer went to a market and purchased a wolf, a goat, and a cabbage. On his way home, the farmer came to the bank of a river and rented a boat. But crossing the river by boat, the farmer could carry only himself and a single one of his purchases: the wolf, the goat, or the cabbage. If left unattended together, the wolf would eat the goat, or the goat would eat the cabbage. The farmer's challenge was to carry himself and his purchases to the far bank of the river, leaving each purchase intact."

Answer the following problem:

1) Represent the WGC problem as a CSP problem.

Variable	{wolf, Goat, Cabbage}
Domain for each variable	{1 (Left),2 (Right)}
Constrains	- Everyone on right side of the river - Wolf in the same side with Goat - Goat in the same side with Cabbages

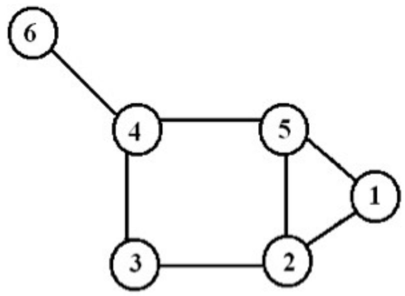
2) Apply backtracking search on your problem for 3 iterations. Show your work.



Waris Damkham Sec 1 6388014

Part II: Local Search

Consider the problem of graph coloring problem with the following graph and setting.



We want to color the above graph using 3 colors (Red Green Blue). Notice that it is impossible.

1) Setup this problem for the hill-climbing search.

Stage	All nodes, Network of graph
Goal	No Neighbor has same color
Action	Selecting the color
Heuristics	Number of available colors

2) Perform the hill-climbing search with your setting for 3 iterations. Show your work.

1 R	1 G	1 B
2 B	2 R	2 G
3 G	3 B	3 R

3) Setup this problem for the GA. You do not need to perform any search for this part.

Algorithm	GA for graph coloring
Data	Adjacency matrix of a graph G
Output	Valid K-coloring of the graph (K = chromatic member of the graph)
Percentage of best fitness	Encountered 90% (individual)

Part III: ML

For the following data examples, answer the following questions.

Training data

#	BMI	BP	Cough	Fever	Disease
1	overweight	low	No	Yes	yes
2	normal	low	Yes	Yes	yes
3	normal	high	No	No	yes
4	overweight	normal	No	No	yes
5	underweight	low	No	No	yes
6	normal	high	Yes	No	yes
7	normal	normal	No	Yes	yes
8	underweight	low	Yes	No	yes
9	underweight	normal	No	No	no
10	overweight	low	Yes	No	no
11	underweight	normal	No	Yes	no
12	underweight	low	Yes	Yes	no
13	overweight	high	Yes	No	no
14	overweight	high	Yes	Yes	no

Test data:

#	BMI	BP	Cough	Fever	Disease
1	overweight	high	No	Yes	?
2	normal	normal	No	No	?
3	underweight	low	Yes	No	?

1) Create a decision tree using the ID3 with the maximum depth of 3. You must show the Info() and the E() of all steps.
2) Extract all rules from your decision tree

3) What are the prediction results from your tree on the test data?
4) What are the prediction results for k-NN using hemming distance with k = 1?
5) What are the prediction results for k-NN using hemming distance with k = 3?
6) What are the prediction results for naïve bayes?

2) What are the pros and cons of evaluating your ML models using a single split vs k-fold cross validation?

3) If you want to know if your current decision tree is in the overfitting or underfitting, how would you check it?

Answer

1. Info $(f^{+}, f^{-}) = f^{-}\lg(f^{+}) - f^{-}\lg(f^{-})$,

   Entropy = $\frac{-P}{P+N}\log 2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log 2\left(\frac{N}{P+N}\right)$

BMI:

Entropy(S) = $\frac{-8}{14}\log 2\left(\frac{8}{14}\right) - \frac{6}{14}\log 2\left(\frac{6}{14}\right)$ = 0.985

S = $[8^{+}, 6^{-}]$

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Normal) = $\frac{-4}{4}\log 2\left(\frac{4}{4}\right) - \frac{0}{4}\log 2\left(\frac{0}{4}\right)$ = 0

S = $[4^{+}, 0^{-}]$

Entropy (Underweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{2}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

I(BMI) = $\frac{P_{Normal}+N_{Normal}}{P+N}$ Entropy (Normal) +

$\frac{P_{Overweight}+N_{Overweight}}{P+N}$ Entropy (Overweight) +

$\frac{P_{Underweight}+N_{Underweight}}{P+N}$ Entropy (Underweight)

= $\frac{4}{14} * 0 + \frac{5}{14} * 0.971 + \frac{5}{14} * 0.971$

= 0.693

Gain (BMI) = 0.985 – 0.693 = 0.292

BP:

Entropy(S) = $\frac{-8}{14}\log 2\left(\frac{8}{14}\right)-\frac{6}{14}\log 2\left(\frac{6}{14}\right)$ = 0.985

S = $[8^{+},6^{-}]$

Entropy (Normal) = $\frac{-2}{4}\log 2\left(\frac{2}{4}\right)-\frac{2}{4}\log 2\left(\frac{2}{4}\right)$ = 1

S = $[2^{+},2^{-}]$

Entropy (Low) = $\frac{-4}{6}\log 2\left(\frac{4}{6}\right)-\frac{2}{6}\log 2\left(\frac{2}{6}\right)$ = 0.918

S = $[4^{+},2^{-}]$

Entropy (High) = $\frac{-2}{4}\log 2\left(\frac{2}{4}\right)-\frac{2}{4}\log 2\left(\frac{2}{4}\right)$ = 1

S = $[2^{+},2]$

I(BP) = $\frac{P_{Normal}+N_{Normal}}{P+N}$ Entropy (Normal) + $\frac{P_{Low}+N_{Low}}{P+N}$ Entropy (Low) + $\frac{P_{High}+N_{High}}{P+N}$ Entropy (High)

= $\frac{4}{14}*1+\frac{6}{14}*0.918+\frac{4}{14}*1$

= 0.693

Gain (BMI) = 0.985 – 0.964 = 0.021

Cough:

Entropy(S) = $\frac{-8}{14}\log 2\left(\frac{8}{14}\right)-\frac{6}{14}\log 2\left(\frac{6}{14}\right)$ = 0.985

S = $[8^{+},6^{-}]$

Entropy (Yes) = $\frac{-3}{7}\log 2\left(\frac{3}{7}\right)-\frac{4}{7}\log 2\left(\frac{4}{7}\right)$ = 0.985

S = $[3^{+},4^{-}]$

Entropy (No) = $\frac{-5}{7}\log 2\left(\frac{5}{7}\right)-\frac{2}{7}\log 2\left(\frac{2}{7}\right)$ = 0.863

S = $[5^{+},2^{-}]$

I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{7}{14}*0.985+\frac{7}{14}*0.863$

= 0.924

Gain (BMI) = 0.985 – 0.924 = 0.061

Fever:

Entropy(S) = $\frac{-8}{14}\log 2\left(\frac{8}{14}\right) - \frac{6}{14}\log 2\left(\frac{6}{14}\right)$ = 0.985

S = $[8^{+}, 6^{-}]$

Entropy (Yes) = $\frac{-3}{6}\log 2\left(\frac{3}{6}\right) - \frac{3}{6}\log 2\left(\frac{3}{3}\right)$ = 1

S = $[3^{+}, 3^{-}]$

Entropy (No) = $\frac{-5}{8}\log 2\left(\frac{5}{8}\right) - \frac{3}{8}\log 2\left(\frac{3}{8}\right)$ = 0.954

S = $[5^{+}, 3^{-}]$

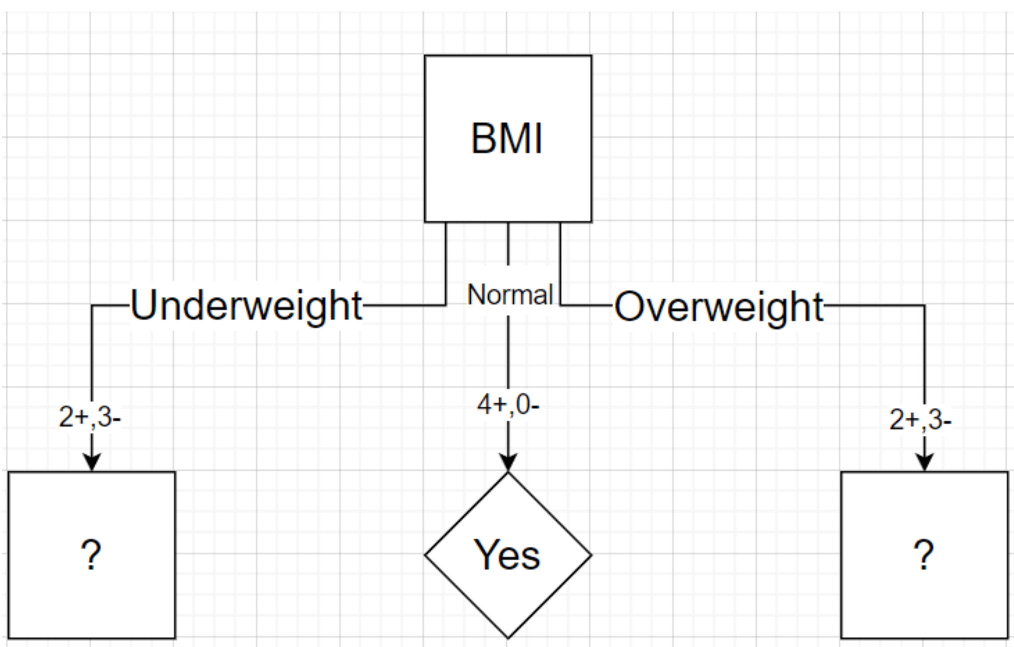
I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{6}{14} * 1 + \frac{8}{14} * 0.954$

= 0.973

Gain (BMI) = 0.985 – 0.973 = 0.012

Attributes	Gain
BMI	0.292
BP	0.021
Cough	0.061
Fever	0.012



Overweight:

BP:

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Normal) = $\frac{-1}{1}\log 2\left(\frac{1}{1}\right) - \frac{0}{1}\log 2\left(\frac{0}{1}\right)$ = 0

S = $[1^{+}, 0^{-}]$

Entropy (Low) = $\frac{-1}{2}\log 2\left(\frac{1}{2}\right) - \frac{1}{2}\log 2\left(\frac{1}{2}\right)$ = 1

S = $[1^{+}, 1^{-}]$

Entropy (High) = $\frac{0}{2}\log 2\left(\frac{0}{2}\right) - \frac{0}{2}\log 2\left(\frac{0}{2}\right)$ = 0

S = $[0^{+}, 2]$

I(BP) = $\frac{P_{Normal}+N_{Normal}}{P+N}$ Entropy (Normal) + $\frac{P_{Low}+N_{Low}}{P+N}$ Entropy (Low) + $\frac{P_{High}+N_{High}}{P+N}$ Entropy (High)

= $\frac{1}{5} * 0 + \frac{2}{5} * 1 + \frac{2}{5} * 0$

= 0.4

Gain (BMI) = 0.971 – 0.4 = 0.571

Cough:

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Yes) = $\frac{0}{3}\log 2\left(\frac{0}{3}\right) - \frac{3}{3}\log 2\left(\frac{3}{3}\right)$ = 0

S = $[0^{+}, 3^{-}]$

Entropy (No) = $\frac{-2}{2}\log 2\left(\frac{2}{2}\right) - \frac{0}{2}\log 2\left(\frac{0}{2}\right)$ = 0

S = $[2^{+}, 0^{-}]$

I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{3}{5} * 0 + \frac{2}{5} * 0$

= 0

Gain (BMI) = 0.971 – 0 = 0.971

Fever:

Entropy (Overweight) = $\frac{-2}{5} \log 2 \left(\frac{2}{5}\right) - \frac{3}{5} \log 2 \left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Yes) = $\frac{-1}{2} \log 2 \left(\frac{1}{2}\right) - \frac{1}{2} \log 2 \left(\frac{1}{2}\right)$ = 1

S = $[1^{+}, 1^{-}]$

Entropy (No) = $\frac{-1}{3} \log 2 \left(\frac{1}{3}\right) - \frac{2}{3} \log 2 \left(\frac{2}{3}\right)$ = 0.918

S = $[1^{+}, 2^{-}]$

I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{2}{5} * 1 + \frac{3}{5} * 0.918$

= 0.950

Gain (BMI) = 0.971 – 0.950 = 0.021

Attributes	Gain
BP	0.571
Cough	0.971
Fever	0.021

Underweight:

BP:

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Normal) = $\frac{-0}{2}\log 2\left(\frac{0}{2}\right) - \frac{2}{2}\log 2\left(\frac{2}{2}\right)$ = 0

S = $[0^{+}, 2^{-}]$

Entropy (Low) = $\frac{-2}{3}\log 2\left(\frac{2}{3}\right) - \frac{1}{3}\log 2\left(\frac{1}{3}\right)$ = 0.918

S = $[2^{+}, 1^{-}]$

Entropy (High) = $\frac{0}{0}\log 2\left(\frac{0}{0}\right) - \frac{0}{0}\log 2\left(\frac{0}{0}\right)$ = 0

S = $[0^{+}, 0]$

I(BP) = $\frac{P_{Normal}+N_{Normal}}{P+N}$ Entropy (Normal) + $\frac{P_{Low}+N_{Low}}{P+N}$ Entropy (Low) + $\frac{P_{High}+N_{High}}{P+N}$ Entropy (High)

= $\frac{2}{5} * 0 + \frac{3}{5} * 0.92 + \frac{0}{5} * 0$

= 0.552

Gain (BMI) = 0.971 – 0.552 = 0.419

Cough:

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right) - \frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Yes) = $\frac{1}{2}\log 2\left(\frac{1}{2}\right) - \frac{1}{2}\log 2\left(\frac{1}{2}\right)$ = 1

S = $[1^{+}, 1^{-}]$

Entropy (No) = $\frac{-1}{3}\log 2\left(\frac{1}{3}\right) - \frac{2}{3}\log 2\left(\frac{2}{3}\right)$ = 0.918

S = $[1^{+}, 2^{-}]$

I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{3}{5} * 0.92 + \frac{2}{5} * 1$

= 0.952

Gain (BMI) = 0.971 – 0.952 = 0.019

Fever:

Entropy (Overweight) = $\frac{-2}{5}\log 2\left(\frac{2}{5}\right)-\frac{3}{5}\log 2\left(\frac{3}{5}\right)$ = 0.971

S = $[2^{+}, 3^{-}]$

Entropy (Yes) = $\frac{0}{2}\log 2\left(\frac{0}{2}\right)-\frac{2}{2}\log 2\left(\frac{2}{2}\right)$ = 0

S = $[0^{+}, 2^{-}]$

Entropy (No) = $\frac{-2}{3}\log 2\left(\frac{2}{3}\right)-\frac{1}{3}\log 2\left(\frac{1}{3}\right)$ = 0.918

S = $[2^{+}, 1^{-}]$

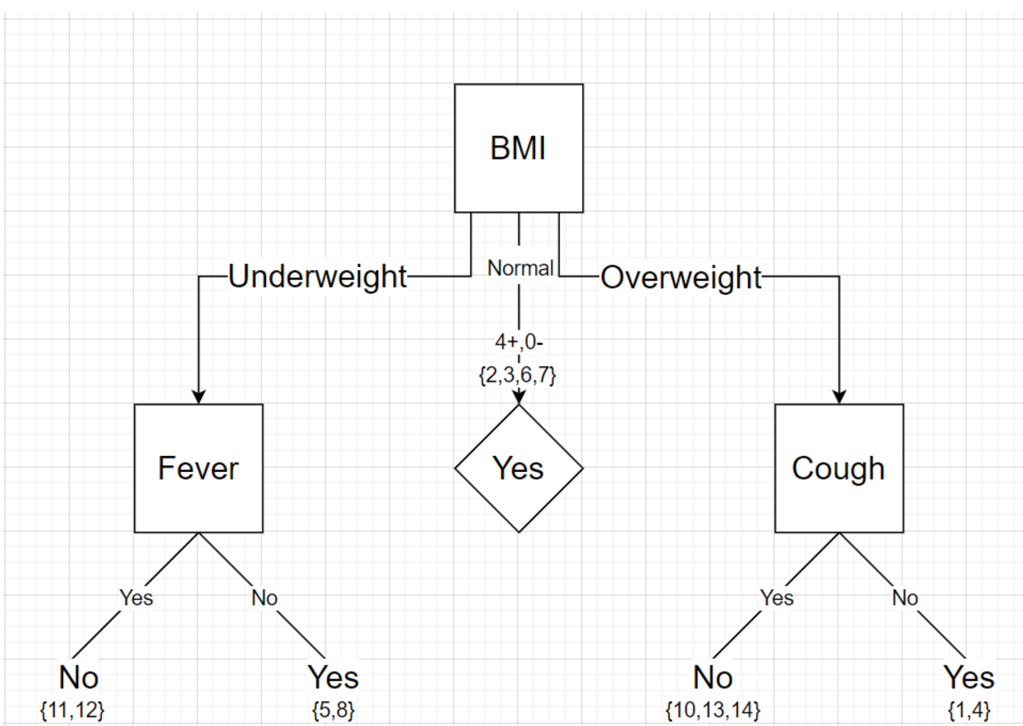
I(BMI) = $\frac{P_{Yes}+N_{Yes}}{P+N}$ Entropy (Yes) + $\frac{P_{No}+N_{No}}{P+N}$ Entropy (No)

= $\frac{2}{5} * 0 + \frac{3}{5} * 0.918$

= 0.550

Gain (BMI) = 0.971 – 0.550 = 0.421

Attributes	Gain
BP	0.419
Cough	0.019
Fever	0.421



2.

If BMI == “Underweight” && Fever == “Yes” -> Disease = “No”

If BMI == “Underweight” && Fever == “No” -> Disease = “Yes”

If BMI == “Norma” -> Disease = “Yes”

If BMI == “Overweight” && Cough == “Yes” -> Disease = “No”

If BMI == “Overweight” && Cough == “No” -> Disease = “Yes”

3. Test Data

	BMI	BP	Cough	Fever	Disease
1	overweight	high	No	Yes	Yes
2	normal	normal	No	No	Yes
3	underweight	low	Yes	No	Yes

4. Test Data When K = 1

	BMI	BP	Cough	Fever	Disease
1	overweight	high	No	Yes	No
2	normal	normal	No	No	Yes
3	underweight	low	Yes	No	Yes

5. Test Data When K = 3

	BMI	BP	Cough	Fever	Disease
1	overweight	high	No	Yes	No
2	normal	normal	No	No	Yes
3	underweight	low	Yes	No	Yes

1 Yes, 2 No

2 Yes, 1 No

2 Yes, 1 No

6. What are the prediction results for naïve bayes?

$$= \frac{p(OverI^{+})p(HighI^{+})p(NoI^{+})p(YesI^{+})p(+)}{p(OverI^{-})p(HighI^{-})p(NoI^{-})p(YesI^{-})p(-)}$$

= 0.9375

$$= \frac{p(NormalI^{+})p(NormalI^{+})p(NoI^{+})p(NoI^{+})p(+)}{p(NormalI^{-})p(NormalI^{-})p(NoI^{-})p(NoI^{-})p(-)}$$

= 0.9375

$$= \frac{p(UnderI^{+})p(LowI^{+})p(YesI^{+})p(NoI^{+})p(+)}{p(UnderI^{-})p(LowI^{-})p(YesI^{-})p(NOI^{-})p(-)}$$

= 0.703

- What are the pros and cons of evaluating your ML models using a single split vs k-fold cross validation?

	Pros	Cons
Single split:	It is better for the small data set.	It may cause some conflict.
K- fold:	Reduced bias	It is not good for small datasets. It may also elicit an appropriate response.

- If you want to know if your current decision tree is in the overfitting or underfitting, how would you check it?
  - If the training and test results are both close, the model has not overfit.
  - If the training result is better than the result, the model has been overfit.
  - If the training and test results are both low, the model has underfit.