Problem 1

i. Since the sum of 16 full joint probabilities in the equals to 1, we have

$P(A \wedge B \wedge C \wedge D) = 1 -$ 0.10-0.01-0.05-0.15-0.02-0.03-0.04-0.05-0.20-0.01-0.01-0.03-0.02-0.04-0.08 =0.16

ii. $P(\neg A|B \wedge C \wedge \neg D) = \dfrac{P(\neg A \wedge B \wedge C \wedge \neg D)}{P(B \wedge C \wedge \neg D)}$ //product rule, Eq. 13.1

numerator = $P(\neg A \wedge B \wedge C \wedge \neg D)$=0.04 //from the full joint prob table

denominator = $P(B \wedge C \wedge \neg D)$

= $P(\neg A \wedge B \wedge C \wedge \neg D) + P(A \wedge B \wedge C \wedge \neg D)$ //via marginalizing over all values of A

= $0.04 + 0.08 = 0.12$

So, the answer is 0.04/0.12=0.33

iii. $P(\neg B|\neg A \wedge D) = \dfrac{P(\neg B \wedge \neg A \wedge D)}{P(\neg A \wedge D)}$ //product rule

numerator = $P(\neg B \wedge \neg A \wedge D)$

$=P(\neg B \wedge \neg A \wedge D \wedge C)+P(\neg B \wedge \neg A \wedge D \wedge C)$ //marginalizing [Eq. 13.4] over C

=0.01+0.15=0.16

denominator = $P(\neg A \wedge D)$ //marginalizing over B,C

$= P(\neg B \wedge \neg A \wedge D \wedge C)+P(\neg B \wedge \neg A \wedge D \wedge \neg C) + P(B \wedge \neg A \wedge D \wedge C)+P(B \wedge \neg A \wedge D \wedge \neg C)$

= 0.01+0.15+0.03+0.05=0.24

So, the answer is 0.16/0.24=0.6667

iv.

iv. $P(\neg B) = P(\neg B \wedge \neg A \wedge \neg C \wedge \neg D) + P(\neg B \wedge \neg A \wedge \neg C \wedge D)$

$+ P(\neg B \wedge \neg A \wedge C \wedge \neg D) + P(\neg B \wedge \neg A \wedge C \wedge D)$

$+ P(\neg B \wedge A \wedge \neg C \wedge \neg D) + P(\neg B \wedge A \wedge \neg C \wedge D)$

$+ P(\neg B \wedge A \wedge C \wedge \neg D) + P(\neg B \wedge A \wedge C \wedge D)$ //marginalizing over A, C, D

= 0.10+0.01+0.05+0.15+0.20+0.01+0.01+0.03=0.56

v. We need to sum over all the cells where the numerator is true, and we also need to sum over the cells where the denominator is true.

$$P(\neg A \vee B \mid C \vee D) = \frac{P((\neg A \vee B) \wedge (C \vee D))}{P(C \vee D)} \quad \text{//product rule}$$

$$\begin{aligned}
\text{numerator} &= P((\neg A \vee B) \wedge (C \vee D)) \\
&= P(\neg A \wedge \neg B \wedge C \wedge \neg D)+P(\neg A \wedge \neg B \wedge \neg C \wedge D) \\
&\quad +P(\neg A \wedge \neg B \wedge C \wedge D)+P(A \wedge B \wedge C \wedge \neg D) \\
&\quad +P(A \wedge B \wedge \neg C \wedge D)+P(A \wedge B \wedge C \wedge D) \\
&\quad +P(\neg A \wedge B \wedge \neg C \wedge D)+P(\neg A \wedge B \wedge C \wedge \neg D)+P(\neg A \wedge B \wedge C \wedge D) \\
&= 0.05+0.01+0.15+0.04+0.08+0.16+0.03+0.04+0.05 \\
&= 0.61
\end{aligned}$$

$$\begin{aligned}
\text{denominator} &= P(C \vee D) \\
&= P(\neg A \wedge \neg B \wedge C \wedge \neg D)+P(\neg A \wedge B \wedge C \wedge \neg D) \\
&\quad + P(A \wedge \neg B \wedge C \wedge \neg D) +P(A \wedge B \wedge C \wedge \neg D) \\
&\quad + P(\neg A \wedge \neg B \wedge C \wedge D) + P(\neg A \wedge B \wedge C \wedge D) \\
&\quad + P(A \wedge \neg B \wedge C \wedge D)+P(A \wedge B \wedge C \wedge D) \\
&\quad +P(\neg A \wedge \neg B \wedge \neg C \wedge D)+P(\neg A \wedge B \wedge \neg C \wedge D) \\
&\quad +P(A \wedge \neg B \wedge \neg C \wedge D)+P(A \wedge B \wedge \neg C \wedge D) \\
&= 0.05+0.04+0.01+0.08+0.15+0.05+0.03+0.16+0.01+0.03+0.01+0.04 \\
&= 0.66
\end{aligned}$$

So, the answer is 0.61/0.66=0.924

Problem 2

(a) Using equation 13.5:

$$P(cc) = P(cc|wh) \times P(wh) + P(cc|{\sim}wh) \times P({\sim}wh)$$

$$0.25 = 0.10 \times 0.75 + P(cc|{\sim}wh) \times 0.25 \quad \text{// since } P({\sim}wh) = 1 - P(wh)$$

So $P(cc|{\sim}wh) = 0.70$ // solving for $P(cc|{\sim}wh)$

(b) $P(wh|cc) = \dfrac{P(wh \wedge cc)}{P(cc)} = \dfrac{P(cc|wh) \times P(wh)}{P(cc)} = \dfrac{0.10 \times 0.75}{0.25} = 0.30$

(c) We wish to compare P(likes|book) (that is, the probability that Margaret likes the book, given the words in it) to P(dislikes|book) and see which is more probable. We'll make the Naive Bayes assumption and suppose that words in books are independent of one another.

To avoid probabilities of zero due to words that we just haven't happened to encounter in the positive or negative training sets, we'll start each counter at 1 instead of 0(by adding another book, which contains all the words once, for both positive and negative training sets). So, the probability of the word 'animal' appearing in a book that Margaret likes is 8/29 (8 times 'animal' appears in the 29 words that appear in books Margaret likes).

Word	animal	mineral	vegetable	see	spot	run	Total Num
Count(like)	7+1	8+1	5+1	3+1	0+1	0+1	29
Count(dislike)	4+1	1+1	6+1	0+1	2+1	9+1	28

$$P(\text{like}|\text{book}) = \frac{P(\text{book}|\text{like}) \times P(\text{like})}{P(\text{book})}$$

$$P(\text{dislike}|\text{book}) = \frac{P(\text{book}|\text{dislike}) \times P(\text{dislike})}{P(\text{book})}$$

$P(\text{book}|\text{like}) = P(\text{see,spot,spot,animal,run}|\text{like})$

$=P(\text{see}|\text{like}) \times P(\text{spot}|\text{like}) \times P(\text{spot}|\text{like}) \times P(\text{animal}|\text{like}) \times P(\text{run}|\text{like})$

$$= \frac{4}{29} \times \frac{1}{29} \times \frac{1}{29} \times \frac{8}{29} \times \frac{1}{29} = 1.56 \times 10^{-6}$$

$P(\text{book}|\text{dislike}) = P(\text{see,spot,spot,animal,run}|\text{dislike})$

$=P(\text{see}|\text{dislike}) \times P(\text{spot}|\text{dislike}) \times P(\text{spot}|\text{dislike}) \times P(\text{animal}|\text{dislike}) \times P(\text{run}|\text{dislike})$

$$= \frac{1}{28} \times \frac{3}{28} \text{v} \frac{3}{28} \times \frac{5}{28} \times \frac{10}{28} =2.61 \times 10^{-5}$$

$$\text{So, } \frac{P(\text{like}|\text{book})}{P(\text{dislike}|\text{book})} = \frac{P(\text{book}|\text{like}) \times P(\text{like})}{P(\text{book}|\text{dislike}) \times P(\text{dislike})} = \frac{1.56 \times 10^{-6} \times (6/17)}{2.61 \times 10^{-5} \times (11/17)} < 1$$

Thus, it is more probable that Margaret does not like the book.

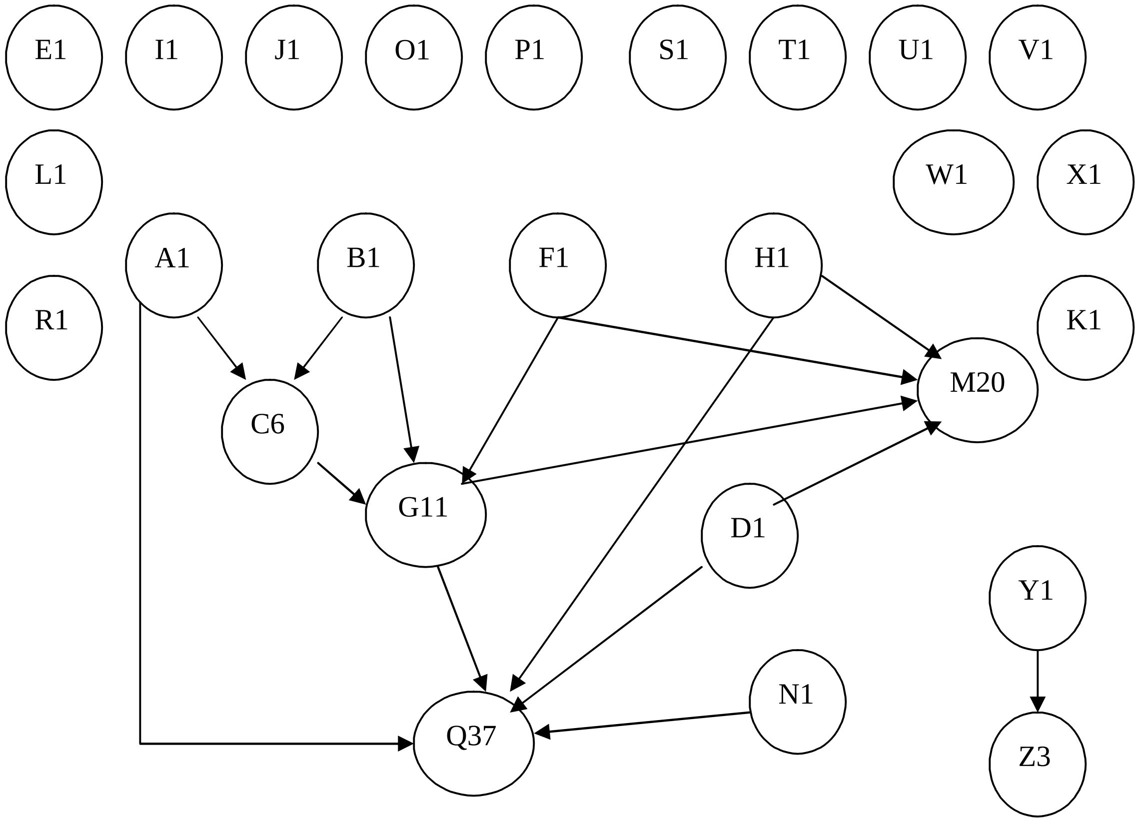
Problem 3

i. It should be all combinations, minus one because we know the probabilities sum to one
2^{26}-1 = 67,108,863 entries (it's also OK if you did not subtract the '1')

ii. 26 entries. We only need to store the prior probabilities of each variable.

iii. Each variable is shown with the memory requirement printed after it. "A1" means variable A needs to store 1 number. Each number requires one number (probability) for each variable combination of their parents, plus one more to store each parent link. That is, memory required = $np + p$, where n is the number of variable values (2 for all variables in this case), and p is the number of parents.

Note that E I J K L O P R S T U V W and X are not connected to anything. This is okay, but they are still part of the system, and are necessary to calculate the probability of an event. They all need 1 number to represent them--their prior probability of being true.



The total memory required is 98 numbers.

Problem 4

i. Just need to use the CPTs and net structure:

$$P(A \wedge \neg B \wedge C \wedge \neg D) = P(A) \times P(C) \times P(\neg B|A) \times P(\neg D|\neg B \wedge C)$$

$$= 0.4 \times 0.7 \times 0.7 \times 0.7 = 0.1372$$

ii. $P(D|\neg A \wedge B \wedge \neg C) = \dfrac{P(D \wedge \neg A \wedge B \wedge \neg C)}{P(\neg A \wedge B \wedge \neg C)}$ //product rule

$$P(D \wedge \neg A \wedge B \wedge \neg C) = P(\neg A) \times P(B \mid \neg A) \times P(\neg C) \times P(D \mid B \wedge \neg C)$$

$$= 0.6 \times 0.9 \times 0.3 \times 0.5 = 0.081$$

$P(\neg A \wedge B \wedge \neg C) = P(D \wedge \neg A \wedge B \wedge \neg C) + P(\neg D \wedge \neg A \wedge B \wedge \neg C)$ //via marginalizing over D

$$= 0.081 + P(\neg A) \times P(B \mid \neg A) \times P(\neg C) \times P(\neg D \mid B \wedge \neg C)$$

$$= 0.081 + 0.6 \times 0.9 \times 0.3 \times 0.5 = 0.162$$

$$P(D|\neg A \wedge B \wedge \neg C) = 0.081 / 0.162 = 0.5$$

iii. $P(A \mid \neg B \wedge C \wedge \neg D) = \dfrac{P(A \wedge \neg B \wedge C \wedge \neg D)}{P(\neg B \wedge C \wedge \neg D)}$ //product rule

$P(A \wedge \neg B \wedge C \wedge \neg D)$ was already calculated in (i)

$P(\neg B \wedge C \wedge \neg D) = P(\neg A \wedge \neg B \wedge C \wedge \neg D) + P(A \wedge \neg B \wedge C \wedge \neg D)$

//via marginalizing over all values of A

$P(\neg A \wedge \neg B \wedge C \wedge \neg D) = P(\neg A) \times P(\neg B \mid \neg A) \times P(C) \times P(\neg D \mid \neg B \wedge C)$

$= 0.6 \times 0.1 \times 0.7 \times 0.7 = 0.0294$

So, $P(A \mid \neg B \wedge C \wedge \neg D) = 0.1372 / (0.1372 + 0.294) = 0.8235$

iv

$P(\neg B \mid A \wedge \neg C) = \dfrac{P(\neg B \wedge A \wedge \neg C)}{P(A \wedge \neg C)}$ //produc rule

$P(\neg B \wedge A \wedge \neg C) = P(\neg B \wedge A \wedge \neg C \wedge \neg D) + P(\neg B \wedge A \wedge \neg C \wedge D)$ //marginalizing over all values of D

$P(A \wedge \neg C) = P(A \wedge \neg B \wedge \neg C \wedge \neg D) + P(A \wedge B \wedge \neg C \wedge \neg D)$

$+ P(A \wedge \neg B \wedge \neg C \wedge D) + P(A \wedge B \wedge \neg C \wedge D)$ //marginalizing over all values of B, D

Now we have everything in terms of Probs with all four variables specified,

so apply the rule - Eq 14.1 - for calculating with a Bayes net and plug in the numbers.

$P(\neg B \wedge A \wedge \neg C \wedge \neg D) = 0.4 \times 0.7 \times 0.3 \times 0.2 = 0.0168$

$P(\neg B \wedge A \wedge \neg C \wedge D) = 0.4 \times 0.7 \times 0.3 \times 0.8 = 0.0672$

$P(A \wedge B \wedge \neg C \wedge \neg D) = 0.4 \times 0.3 \times 0.3 \times 0.5 = 0.018$

$P(A \wedge B \wedge \neg C \wedge D) = 0.4 \times 0.3 \times 0.3 \times 0.5 = 0.018$

$P(\neg B \wedge A \wedge \neg C) = 0.084$

$P(A \wedge \neg C) = 0.12$

$P(\neg B \mid A \wedge \neg C) = 0.084 / 0.12 = 0.7$

v. $P(\neg B) = P(\neg B \times \neg A \times \neg C \times \neg D) + P(\neg B \times \neg A \times \neg C \times D)$ //marginalizing over all values of A, C, D

$+ P(\neg B \times \sim A \times C \times \neg D) + P(\neg B \times \neg A \times C \times D)$

$+ P(\neg B \times A \times \neg C \times \neg D) + P(\neg B \times A \times \neg C \times D)$

$+ P(\neg B \times A \times C \times \neg D) + P(\neg B \times A \times C \times D)$

Again, we have all the probs on the right-hand side in terms of cases

where the values of all four variables are specified.

$P(\neg B \times \neg A \times \neg C \times \neg D)=0.0036$; $P(\neg B \times \neg A \times \neg C \times D)=0.0144$

$P(\neg B \times \sim A \times C \times \neg D)=0.0294$; $P(\neg B \times \neg A \times C \times D)=0.0126$

$P(\neg B \times A \times \neg C \times \neg D)=0.0168$; $P(\neg B \times A \times \neg C \times D)=0.0672$

$P(\neg B \times A \times C \times \neg D)=0.1372$; $P(\neg B \times A \times C \times D)=0.0588$

$P(\neg B) = 0.34$

Problem 5

One possible similarity function is to count the number of mismatched feature values (this is called the *Hamming distance*). Using this measure, the distance between the test example and each training example is:

```
Dist(ex1, test) = 2 +
Dist(ex2, test) = 1 -
Dist(ex3, test) = 1 +
Dist(ex4, test) = 1 +
Dist(ex5, test) = 3 -
Dist(ex6, test) = 2 +
```

The nearest three neighbors are ex2, ex3, ex4, of which the majority class label is "+". So the test should be classified as "+".