Final for 6.034 Spring 2010

Name:

20	20	20	20	20

Good luck!

Question #1 20 points

Indicate whether the following statement is true and briefly justify your answer.

(i) (2 points) If we add more training data, the number of support vectors will always increase.

(ii) (2 points) For a given training set, support vectors are invariant to the kernel choice.

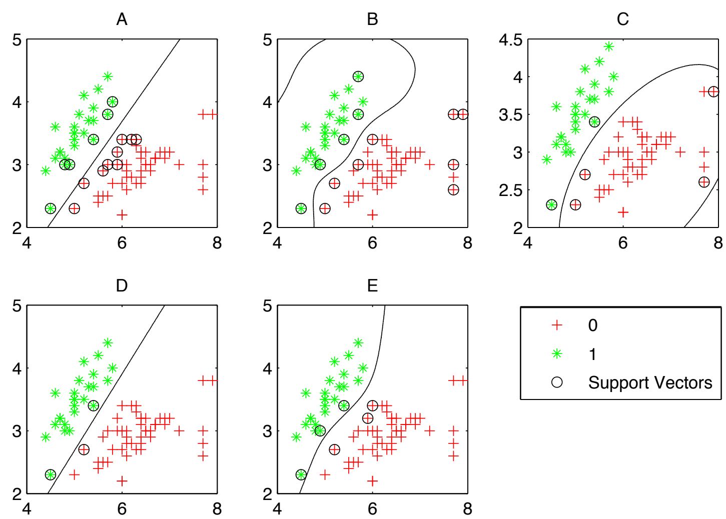
(iii) (2 points) There is a distribution over five variables that can be represented by any five node Bayesian network topology.

(iv) (2 points) The largest factor created by the variable elimination algorithm is no bigger than the largest initial factor.

(v) (2 points) The variable elimination algorithm can require time exponential in the size of the Bayesian network.

(vi) (2 points) Observing an additional node (i.e. making the node an evidence node) in a Bayesian network can only increase the number of variable pairs which are conditionally independent.

Each of the five plots (A to E) below shows the result of training a soft-margin SVM on a two-class data set (labels are 0 and 1) with different kernel functions and regularization parameter C.



(vii) (8 points) For each plot, find the best matching set of learning parameters (i.e. kernel function and C). Each set of parameters can only be used once.

Plot	Kernel	C
	Linear	1
	Linear	1000
	Polynomial (degree 2)	1000
	Polynomial (degree 5)	1000
	Gaussian/RBF	1000

Question #2 20 points

Here, you plan to use a naive Bayesian classifier to build a personal computer troubleshooting system for novices to decide whether the hardware is *bad* or *ok*. However, for simplicity, this decision must be based only on three noisy boolean (*true*/*false*) observations:

- N: Computer makes loud (n)oises
- F: Computer (f)reezes frequently
- A: (A)pplications run slowly

Cause: $H \in \{bad, ok\}$
Evidence: $N, F, A \in \{true, false\}$
You goal is to build a classifier that models $P(H|N, F, A)$.

(a) (4 points) Under the naive Bayes assumption, write an expression for $P(H|N, F, A)$ in terms of *only* probabilities of the forms $P(H), P(N|H), P(F|H)$, and $P(A|H)$

(b) (6 points) The company has records for several recent user support cases:

H	N	F	A
ok	false	false	true
ok	false	true	false
ok	false	true	false
ok	true	false	false
ok	false	false	false
bad	true	false	false
bad	true	true	false
bad	false	true	true

Given this data, fill the tables below with the parameter estimates for the parameters of the naive Bayesian classifier. Use Laplace correction (aka "add-one" smoothing) for feature estimates $P(N|H), P(F|H)$, and $P(A|H)$ but *not* for class prior estimates $P(H)$.

H	P(H)
bad	3/8
ok	/

N	H	P(N|H)
true	bad	/
false	bad	/
true	ok	/
false	ok	/

F	H	P(F|H)
true	bad	/
false	bad	/
true	ok	/
false	ok	/

A	H	P(A|H)
true	bad	/
false	bad	/
true	ok	/
false	ok	/

Suppose you are given an observation of a new case: N=false, F=true, A=true.

(c) (4 points) Calculate the joint probability $P(H, N = false, F = true, A = true)$. Please leave the probabilities in fractional form.

H	P(H, N=false, F=true, A=true)
ok	
bad	

(d) (2 points) State the prediction of your naive Bayes classifier, i.e. bad or ok.

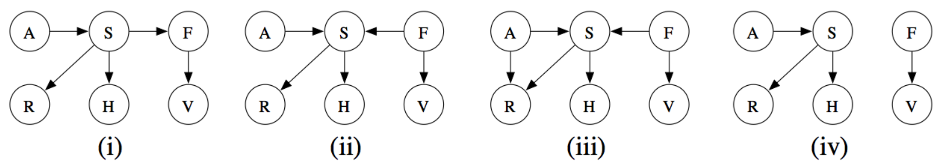
(e) (2 points) Calculate the posterior probability P(H=ok|N=false, F=true, A=true). (Please leave it in fractional form.)

(f) (2 points) Suppose you use hard-margin SVMs instead. Give a kernel function that can separate the training data. If no such kernel exists, briefly explain why.

Question #3 20 points

Assume there are two types of conditions: (S)inus congestion and (F)lu. Sinus congestion is caused by (A)llergy or the flu.

There are three observed symptoms for these conditions: (H)eadache, (R)unny nose, and fe(V)er. Runny nose and headaches are directly caused by sinus congestion (only), while fever comes from having the flu (only). For example, allergies only cause runny noses indirectly. Assume each variable is binary (i.e. true/false).



(a) (5 points) Consider the four Bayesian networks shown. Indicate which one models the domain (as described above) best.

Best network : i / ii / iii / iv

(b) (5 points) Assume we wanted to remove the Sinus congestion (S) node. Using the least number of edges, draw a Bayesian network over the remaining variables which can encode the original model's marginal distribution over the remaining variables.

The following samples were drawn from the correct Bayesian network using prior sampling:

Sample #	A	S	R	H	F	V
1	true	true	true	false	false	false
2	true	true	false	true	true	false
3	true	false	false	false	false	false
4	true	false	false	true	true	false
5	true	true	false	true	false	false

(c) (2 points) Give the sample estimate of $P(F = true)$ or state why it cannot be computed.

(d) (2 points) Give the sample estimate of $P(F = true|H = true)$ or state why it cannot be computed.

(e) (2 points) Give the sample estimate of $P(F = true|V = true)$ or state why it cannot be computed.

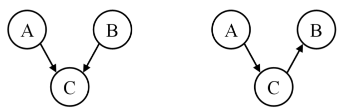
(f) (4 points) For rejection sampling in general (not necessarily on these samples), which query will require more samples to compute a certain degree of accuracy, $P(F|H)$ or $P(F|H, A)$? Briefly justify.

Question #4 20 points

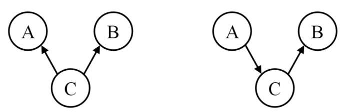
Consider the following pairs of Bayesian networks. **If the two networks have identical conditional independence properties**, indicate *same* and write one of their shared independence (or *none* if they assert none.) **Otherwise**, indicate *different* and write an independence property that one has but the other. For example, in the following case you would answer as shown:

different: right has $A \perp B$

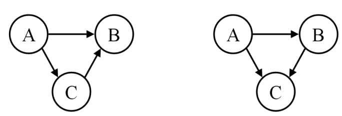
(a) (3 points)



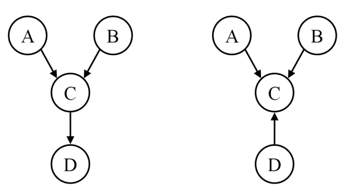
(b) (3 points)



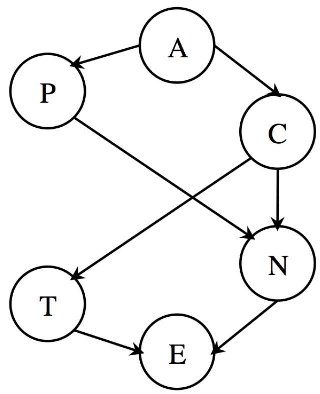
(c) (3 points)



(d) (3 points)



Consider the following Bayesian network.

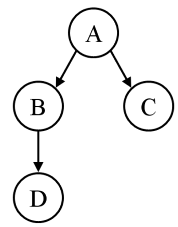


(e) (8 points) For each the following conditional independence (CI) property, indicate whether it is true, false, or unknown.

CI property	Assertion
$N \perp T$	True / False / Unknown
$P \perp E \mid N$	True / False / Unknown
$P \perp N \mid \{A,C\}$	True / False / Unknown
$E \perp A \mid \{C,N\}$	True / False / Unknown

Question #5 — 20 points

The next parts involve computing various quantities in the network below. These questions are designed so that they can be answered with minimum computation. If you find yourself doing copious amount of computation for each part, step back and consider whether there is a simpler way to deduce the answer.



A	P(A)
true	0.1
false	0.9

B	A	P(B|A)
true	true	0.5
false	true	0.5
true	false	0.8
false	false	0.2

C	A	P(C|A)
true	true	0.4
false	true	0.6
true	false	0.7
false	false	0.3

D	B	P(D|B)
true	true	0.9
false	true	0.1
true	false	0.2
false	false	0.8

(a) (2 points) P(A=true, B=false, C=true, D=false)

(b) (2 points) P(A=true, B=true)

(c) (2 points) P(B=true)

(d) (2 points) P(A=true|B=true)

(e) (2 points) P(D=true|A=true)

(f) (2 points) P(D=true|A=true,C=true)

Consider computing the following distributions in the above network using various methods:

(i) P(A|B=true,C=true,D=true)

(ii) P(C|D=true)

(iii) P(D|A=true)

(iv) P(D)

(g) (2 points) Which query is least expensive using inference by enumeration? Briefly justify.

(h) (2 points) Which query is most improved by using likelihood weighting instead of rejection sampling (in terms of number of samples required)? Briefly justify.

(i) (2 points) When computing P(D) with variable elimination, what factor(s) is/are eliminated by first eliminating variable A from the active list?

(j) (2 points) When computing P(D) with variable elimination, show the new factor (in tabular form) that is created by first eliminating variable A from the active list.