Question 1 - 15 points

1a (5 points). Function P_1 is a function defined on a set of samples S = {A, B, C, D}. P_1 is defined as shown below. Is there a value for y that makes P_1 a valid **probability function**? If yes, what is that value? Justify your answer.

$P_1(A) = 10*y$
$P_1(B) = 20*y$
$P_1(C) = 30*y$
$P_1(D) = 20*y$

1b (5 points). Function P_2 is a function defined on the set of real numbers. P_2 is defined as shown below. Is there a value for y that makes P_2 a valid **probability density function**? If yes, what is that value? Justify your answer.

$P_2(x) = 0$ if $x < 100$
$P_2(x) = 6*y$ if $100 <= x <= 110$
$P_2(x) = 4*y$ if $110 <= x <= 130$
$P_2(x) = 0$ if $x > 130$.

1c (5 points). Function P_3 is a function defined on the set of real numbers. P_3 is defined as shown below. Is there a value for y that makes P_3 a valid **probability density function**? If yes, what is that value? Justify your answer.

$P_3(x) = 0$ if $x < 0$
$P_3(x) = 7*y$ if $0 <= x <= 10$
$P_3(x) = 3*y$ if $x > 10$

Question 2 – 10 points

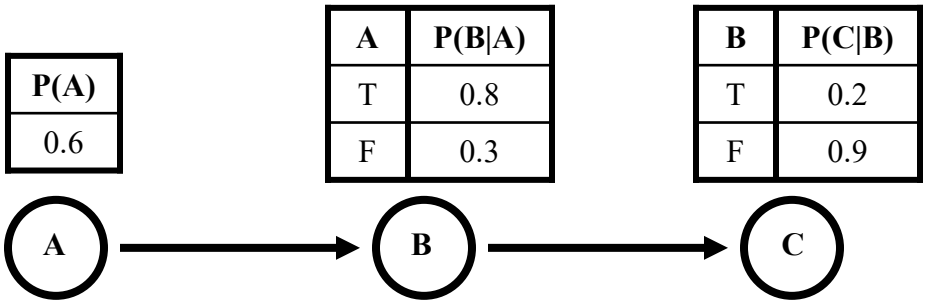
Age of owner	Car	Minivan	SUV
under 30	0.15	0.05	0.1
between 30 and 50	0.1	0.15	0.1
over 50	0.15	0.15	0.05

The above joint distribution table shows the probability of combinations of vehicle types and vehicle owners. For example, the probability that a vehicle is an SUV and is owned by a person over 50 years old is 0.05. Using that table:

2a. (5 points) Determine the probability that a vehicle owner is under 30 years old.

2b. (5 points) Determine P(vehicle type = Minivan | age of owner is under 30)

Question 3 - 10 points



3a. (5 points) Given the above Bayesian network, compute P((B=true) AND (C=false)). You do not have to carry out numerical calculations, but you have to write an expression that fully specifies the answer numerically.

3b. (5 points) In the above Bayesian network, compute P(B). You do not have to carry out numerical calculations, but you have to write an expression that fully specifies the answer numerically.

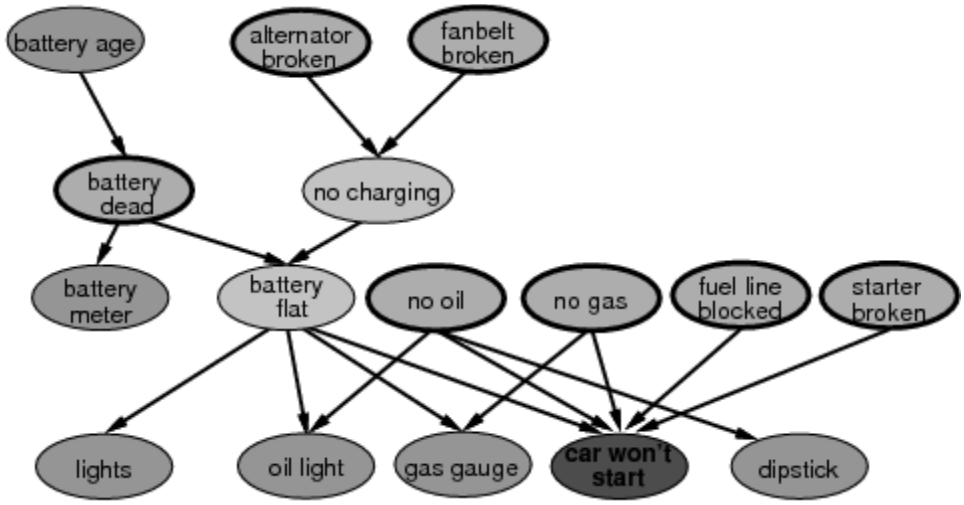
Question 4 = 15 points

4a (5 points). You have a Bayesian network of N nodes. Each node corresponds to a Boolean random variable. Each node has a maximum of 3 parents. How many numbers would you need to specify at most, in order to fully specify the probability distribution modeled by this Bayesian network? In other words, what is the maximum number of values you need to store (for the entire network) in order to full specify the probability table for each node? Justify your answer.

4b (5 points). What is the time complexity of doing inference in the Bayesian network of question 4a?

4c (5 points). Suppose that you want to model the same probability distribution as in question **4a**, but using a joint distribution table. How many values do you need to specify in that case?

Question 5 – 10 points



5a (5 points). In the above Bayesian network, is the battery-flat event conditionally independent of the fanbelt-broken event given a value for the no-charging event? Justify your answer.

5b (5 points). In the above Bayesian network, is the battery-flat event conditionally independent of the no-oil event given a value for the oil-light event? Justify your answer.

Question 6 - 15 points

6a (5 points) Suppose that a decision tree is trained on 1000 training examples, and achieves 90% accuracy on the training examples. What is the smallest and largest accuracy that this decision tree can possibly achieve on a test set of 1000 examples? Justify your answer. You can assume there are only two classes.

6b (5 points). Suppose that a decision tree is trained on 1000 training examples, and achieves 80% accuracy on the training examples. What is the smallest and largest possible value for the entropy at a leaf node of this decision tree? Remember that entropy is measured on the training set. You can assume there are only two classes.

6c (5 points). Suppose that a decision tree is trained on 1000 training examples, and achieves 100% accuracy on the training examples. What is the smallest and largest possible value for the entropy at a leaf node of this decision tree? Again, remember that entropy is measured on the training set. You can assume there are only two classes.

Question 7 - 5 points

We want to build a decision tree that determines whether a new laptop is going to break down or not during its first week. This decision tree is trained on 200 training examples (i.e., 200 cases of new laptops). The only thing that we know about each training example is the operating system that the laptop was running. In particular:

100 laptops in the training set crashed within their first week.
70 of those laptops were running operating system AA.
20 of those laptops were running operating system BB.
10 of those laptops were running operating system CC.

100 laptops in the training set did not crash within their first week.
10 of those laptops were running operating system AA.
20 of those laptops were running operating system BB.
70 of those laptops were running operating system CC.

Determine the entropy gain of choosing, at the root node, the predicate (Operating System = CC) as the test to apply at that node. You do not have to carry out numerical calculations, but you have to write an expression that fully specifies the answer numerically.

Question 8 - 10 points

8a (5 points). Design a perceptron that takes (in addition to the bias input) three inputs x, y, z, and outputs: 1 if x-2y-5 >= z, and 0 otherwise. Fully specify bias input and all weights. Your activation function g should return 1 if the weighted sum of inputs (including the bias input) is greater than or equal to 0, and g should return 0 otherwise.

8b (5 points). Design a neural network that takes X and Y as inputs, outputs 0 if X/Y > 5, and outputs 1 if X/Y <= 5. Fully specify bias input and all weights. Your activation function g should return 1 if the weighted sum of inputs (including the bias input) is greater than or equal to 0, and g should return 0 otherwise.

Question 9 - 10 points

Design a neural network that takes X and Y as inputs, and outputs 1 if X - 3 = 2Y, and 0 otherwise. Fully specify bias input and all weights. Your activation function g should return 1 if the weighted sum of inputs (including the bias input) is greater than or equal to 0, and g should return 0 otherwise. If you are using a perceptron whose weights are fully specified in the textbook, just give the name of the perceptron, you do not have to specify those weights.