HONOR SYSTEM: This examination is strictly individual. You are not allowed to talk, discuss, exchange solutions, etc., with other fellow students. Furthermore, you are only allowed to use the book, slides and your class notes. You should not use Internet. You may only ask questions to the class instructor. Any violation of the honor system, or any of the ethic regulations, will be immediately reported according to George Mason University honor court. The full score for the exam is 50 out of 58.

1.		5
2.		5
3.		5
4.		5
5.		5
6.		5
7.		8
8.		5
9.		5
10.		4
11.		6
total		58

1. Suppose that we have a least squares regression problem with n features and one of the features x_j takes only two values -1 and 1. Further whenever $x_j = -1$ the output y is positive and whenever $x_j = +1$ the output is negative. Is it possible for the least squares coefficient θ_j corresponding to the features x_j to be positive ? Assume that θ is chosen by minimizing $\sum_{i=1}^{m}(y^{(i)} - \sum_j \theta_j x_j^{(i)})^2$.

2. Consider the problem of separating N data points into positive and negative examples using linear separator. This can always be done for $N = 2$ and points lying in 1D line $d = 1$.

   a) Show/Argue if/when it can be done for $N = 3$ and $d = 2$, where points line in the plane.

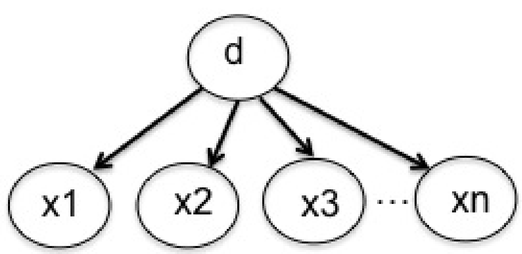
   b) Show/Argue if/when it can be done for $N = 4$ and $d = 3$, where points line in the space of 3D.

3. Consider logical function NAND. Is this function linearly separable ? Can a multilayer neural network learn such function ? If the answer to the previous question is YES draw an example of such network.

4. Discuss how can a Support Vector Machine Learn XOR function. It will be convenient to use values of 1 and -1 instead of 1 and 0 for the inputs and outputs. Consider mapping the input space to the following two dimensions x_1 and x_1x_2. Draw four points in this space and their separator. Draw the separating line back in Euclidean space.

5. A doctor says that a infant who predominantly turns the head to the right will be right-handded and the one who turns the head to the left will be left-handed. The newborn baby girl turned her head to the left. Given that 90% of the population is right-handed, what is the probability of baby being right handed if the test is 90% accurate ? What if it is 80% accurate ?

6. Naive Bayes classifier. Common technique for classification is the Naive Bayes classifier . It uses the following Bayes network to do the classification. Suppose that we want to classify each instance into two classes d_0



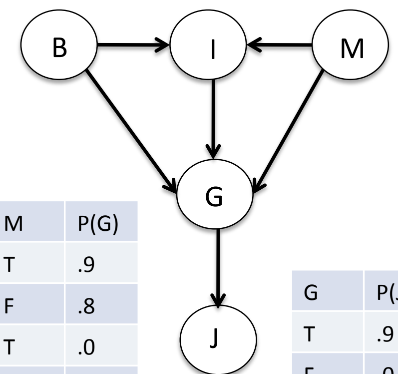
and d_1 and the features values are binary. Write down the formula for deciding whether a new instance $X_1, \ldots X_n$ belongs to d_1. The formula should be written in terms of CPD entries in the network $P(d_0)$ and $P(d_1)$ and $P(X_i|d_0)$ or $P(X_i|d_1)$.

7. A simple Bayes net with boolean variables $B = BrokeElectionRule, I = Indicted, M = PoliticalyMotivatedProsecutor, G = FoundGuilty, J = Jailed.$

B	M	P(I)
T	T	.9
T	F	.5
F	T	.5
F	F	.1

P(B)
0.9

P(M)
.1



B	I	M	P(G)
T	T	T	.9
T	T	F	.8
T	F	T	.0
T	F	F	.0
F	T	T	.2
F	T	F	.1
F	F	T	.0
F	F	F	.0

G	P(J)
T	.9
F	.0

a) Which of the statements below are asserted by the network structure ?

$$P(B, I, M) = P(B)P(I)P(M)$$

$$P(J|G) = P(J|G, I)$$

$$P(M|G, B, I) = P(M|G, B, I, J)$$

b) Calculate the values of $P(b, i, \neg m, g, j)$.

c) Calculate the probability that someone goes to jail given that they broke the law, have been indicted and face politically motivated prosecutor.

d) Suppose we want to add the variable $P = PresidentialPardon$ to the network. Draw the network and explain any links you add.

8. Draw an example of second-order Markov process and show how it can be rewritten as a first order Markov Process with augmented set of state variables. Can this be done without increasing the number of parameters needed to specify the transition model ?

9. The problem of robot localization given the map is the most common example where temporal models of uncertainty are being used and is typically approached using various methods for representing and propagating uncertainty; two most commonly used are Kalman Filter and Particle filter.

   a) Describe how is the belief (uncertainty) represented in each of these two methods.

   b) Consider now the robot kidnapping problem and discuss which of these two methods would be more suitable for continuing to track robots position and why ?

10. Short answers, please provide justification.

    True/False: The complexity of variable elimination algorithm for inference in Bayesian Nets does not depend on variable ordering.

    True/False: A* Algorithm cannot be used in robotics because that states, actions and percepts are continuous.

    Suppose that we use least squares linear regression to fit a hypothesis to a training set. We then test the learned function against different set of data and discover that the test error is much larger that the training error. We would like to reduce the test error.

    True/False: Is is likely that training on more data will help significantly.

    True/False: It is possible the our test error would improve by removing some features from the data.

11. Markov Decision Processes (10pt)

Suppose that you have a MDP with reward function $R(s)$, transition probabilities $P_{sa}(s')$ and discount factor $0 \leq \gamma < 1$. We are also given a biased count which lands *Tails* with probability α and *Heads* with probability $(1 - \alpha)$. Suppose you have a policy π that behaves in a following way: for a given state s, the policy first tosses a coin and if the coin lands *Heads* is executes the optimal policy π^*; if the lands *Tails* it executes some other fixed policy $\hat{\pi}$.

Express the transition probability $P_{s\pi(s)}(s')$ in terms α, $P_{s\pi^*(s)}(s')$ and $P_{s\hat{\pi}(s)}(s')$.

For any state s, show that the following statement relating the value function V^π for policy π with the value function V^{π^*} for the optimal policy π holds:

$$V^\pi(s)-V^{\pi^*}(s) = \gamma \sum_{s' \in S} [P_{s\pi(s)}(s')(V^\pi(s')-V^{\pi^*}(s'))+(P_{s\pi(s)}(s')-P_{s\pi^*(s)}(s'))V^{\pi^*}(s')]$$