CS 726: Quiz 1

Jan 24, 2019. 3:40 to 4:50pm

Roll: ____________________

Name: ____________________

This quiz is open notes.

---

1. Let X and Y be two binary random variables, each of which takes +1 with probability 1/2, and -1 with probability 1/2. Calculate the expected value $E(X.Y)$ under the following two cases.

   (a) X and Y are correlated ..1
   The expectation will be a positive quantity if the correlation is positive and negative if the two are negatively correlated. The exact value cannot be determined based on the information given in this question.

   (b) X and Y are independent 0 ..1

2. Let X and Y be two binary random variables each of which takes value either 0 or 1 and the probability of 1 for each is p. What is the probability that $X = 1$ given that $X + Y < 2$, that is $\Pr(X = 1|X + Y < 2)$? ..2 $p(1-p)/(2p(1-p)+p^2)$

3. Let X and Y be two continuous uniformly distributed random variables between 0 and 1. Calculate the probability $\Pr(X + Y < 1|X > 1/2)$ ..2 $1/4$

4. In the following multiple choice questions, wrong answers will carry negative marks of 1, so do not do random guesses. Let $f(x)$ be a continuous differential function over a real number x. Let $f'(x)$ denote its derivative and $f''(x)$ denotes its second derivation.

   (a) Select all that are true. Assume $\epsilon > 0$ and very small.
      i. $f(x - \epsilon\text{sign}(f'(x))) \leq f(x)$.
      ii. $f(x - \epsilon\text{sign}(f'(x))) \leq f(x)$ only when $f(x) < 0$
      iii. $f(x - \epsilon\text{sign}(f'(x))) \leq f(x)$ only when $f(x)$ is convex.
      iv. When $f(x)$ is convex, $f(x - \epsilon f'(x)) \leq f(x)$.

   ..2 a,d

   (b) A point x^* is a local maxima of $f(x)$ when
      i. $f'(x^*) = 0, \quad f''(x^*) = 0$
      ii. $f'(x^*) = 0, \quad f''(x^*) > 0$
      iii. $f'(x^*) = 0, \quad f''(x^*) < 0$
      iv. $x^* f'(x^*) > 0, \quad f''(x^*) > 0$
      v. $f'(x^*) = 0, \quad f''(x^*) < f'(x^*)$

   ..1 c,e

5. Consider a training dataset for one dimensional data D with just two instances: One positive labeled instance at $x = -1$ and another negative labeled instance at $x = 1$. Assume we are training a logistic regression classifier that predicts a class label based on wx. Which the training objectives for learning w via maximum likelihood in terms of w.

..2 $\min_w 2\log(1 + e^w)$

6. Consider a Bayesian network where the nodes $X_1 \dots X_n$ are numbered topologically i.e., if X_i is a parent of X_j then $i < j$.

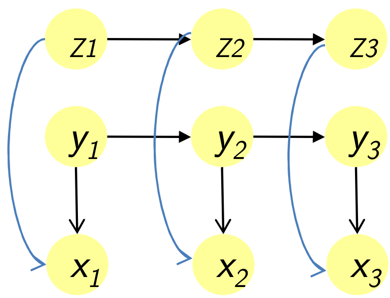
   (a) If X_i has no parents than $X_i \perp\!\!\!\perp X_j$ for any $j < i$. State true or false with a brief justification. ..1 True. Just use the local-CI of BN on X_i and j has to be a non-descendant because of the topological order.

   (b) We can write $P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i))$ as the product of n CPDs (Conditional Probability Distributions). Write the probability of $P(X_1, \dots, X_{n-1})$ in terms of the fewest possible CPDs. ..1 We can just drop the last term and write as $\prod_{i=1}^{n-1} P(X_i|Pa(X_i))$

   (c) If $ND(X_i), D(X_i)$ denotes non-descendants, and descendants of X_i write the simplest possible expression for $P(X_i, ND(X_i), Pa(X_i))$? ..2 $\prod_{j\in\{i,ND(i),Pa(i)\}} P(X_j|Pa(X_j))$

7. Coin 1 has a probability 0.7 of showing heads, coin 2 has a probability of 0.2 of showing heads. A user selects coin 1 with probability 0.3 and coin 2 with probability 0.7, tosses the selected coin and sees the outcome as head. The probability that coin 1 was chosen is ..2 Apply Bayes rule: $\Pr(C_1|H) = \frac{\Pr(H|C_1)\Pr(C_1)}{\Pr(H|C_1)\Pr(C_1)+\Pr(H|C_2)\Pr(C_2)} = \frac{0.7*0.3}{0.7*0.3+0.2*0.7}$

8. Consider the parallel HMM below.



   (a) List a *minimal* set of variables Z for each of the following variable pairs that makes them d-separable.

      i. z_2 and y_2 ..1 empty.

      ii. x_1 and x_3 ..1 any single node from z chain and any single node from y chain.

      iii. x_1 and y_2 ..1

   (b) Draw another Bayesian network that satisfies the same conditional independencies as the one drawn above but comprises of a different edge orientation. ..3

9. Let X, Y, Z be three sets of random variables. Does $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Z$ always imply that $X \perp\!\!\!\perp \{Y, Z\}$. If yes, give a short proof. If no, give a counter example. ..3 Let $X = Y \; xor \; Z$. Knowing value of X does not tell us anything about Y or Z separately but jointly they get constrained.

10. Let matrix $X = [x_1 \quad x_2], \Sigma = \left[\begin{smallmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{smallmatrix}\right]$. Assume none of the elements are zero in X, Σ. Let X^T denote the transpose of vector X. For what values of Σ can you write $X\Sigma X^T$ as $(x_1 - x_2)^2$? ..2 $\Sigma = \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$

11. Using the above definitions of X and Σ, what property should Σ satisfy for $X\Sigma X^T$ to be always positive? ..2 Σ has to be positive definite, that is, diagonal elements must be positive and $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} > 0$

Total: 30