CS 261: Deep Generative Models
Quiz 1 Solutions

Available: 01/23/2024; **Due Date:** 23:59 PM PST, 01/26/2024

General Instructions:

- The quiz contains 10 multiple choice questions. You have 1 hour to finish it. Once submitted, you cannot re-take the quiz.
- The syllabus for this quiz are Discussion 1 (basic probability review) and Lecture 2 (Autoregressive Models, everything before RNNs).
- You are allowed to consult is lecture slides and discussion notes, which you can download in advance and refer to if helpful. No other online or offline resource is permitted.
- The quiz is open till 11:59pm on Friday, Jan 27 2024. There are no late submissions allowed.
- Please follow the UCLA honor code. Any evidence of sharing questions and answers relating to the quiz with other students will lead to an immediate F grade. You are also barred from posting any questions relating to the quizzes on Campuswire until the deadline for submitting the quiz has passed.

1. Consider a collection of n discrete random variables $\{X_i\}_{i=1}^n$, where the number of outcomes for X_i is $|\text{val}(X_i)| = k_i$. Under the full independence assumption (i.e., every variable is independent of every other variable), what is the total number of parameters needed to describe the joint distribution over $(X_1, \dots, X_n)$?

   (a) n
   (b) $(\prod_{i=1}^n k_i) - 1$
   (c) $(\sum_{i=1}^n k_i) - 1$
   (d) $\sum_{i=1}^n (k_i - 1)$

   D. There are $\prod_{i=1}^n k_i$ unique configurations. With full independence assumption, the number of independent parameters needed is $\sum_{i=1}^n (k_i - 1)$.

2. Consider a collection of n discrete random variables $\{X_i\}_{i=1}^n$, where the number of outcomes for X_i is $|\text{val}(X_i)| = k_i$. Without any (conditional) independence assumptions, what is the total number of parameters needed to describe the joint distribution over $(X_1, \dots, X_n)$?

   (a) n
   (b) $(\prod_{i=1}^n k_i) - 1$
   (c) $(\sum_{i=1}^n k_i) - 1$
   (d) $\sum_{i=1}^n (k_i - 1)$

   B. There are $\prod_{i=1}^n k_i$ unique configurations. Without independence assumptions, the number of independent parameters needed is $(\prod_{i=1}^n k_i) - 1$.

3. Which of the following signifies a valid factorization for an autoregressive generative model over an input $\mathbf{x} \in \mathbb{R}^3$?

(a) $p(\mathbf{x}) = p(x_1|x_2, x_3)p(x_2)p(x_3|x_2)$
(b) $p(\mathbf{x}) = p(x_1|x_2, x_3)p(x_2|x_1, x_3)p(x_3|x_1, x_2)$
(c) $p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)$
(d) $p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2)$

A. By chain rule, and from the definition of an autoregressive model, we assume no conditional independence structure.

4. Consider a NADE generative model over a 3D input $\mathbf{x} \in \mathbb{R}^3$. Each $x_i \in \{0, 1\}$ is binary. The first marginal $p(x_1)$ is simply a Bernoulli distribution requiring 1 parameter. Each conditional $p(x_i|x_{<i})$ for $i > 1$ is parameterized via a 1 hidden layer neural network of dimensionality 10. Assume no bias parameters for the hidden or the output layers. With **no sharing** of parameters across the weight matrices for the two conditionals, what is the total number of learnable parameters for NADE?

   (a) 21
   (b) 31
   (c) 41
   (d) 51

   D. 51. 1 (for x_1) + 10 (hidden layer for x_2) + 20 (hidden layer for x_3) + 2x10 (output layer for x_2, x_3).

5. Consider a NADE generative model over a 3D input $\mathbf{x} \in \mathbb{R}^3$. Each $x_i \in \{0, 1\}$ is binary. The first marginal $p(x_1)$ is simply a Bernoulli distribution requiring 1 parameter. Each conditional $p(x_i|x_{<i})$ for $i > 1$ is parameterized via a 1 hidden layer neural network of dimensionality 10. Assume no bias parameters for the hidden or the output layers. With **sharing** of parameters across the weight matrix, what is the total number of learnable parameters for NADE?

   (a) 21
   (b) 31
   (c) 41
   (d) 51

   C. 41. 1 (for x_1) + 10 (hidden layer for x_2) + 10 (hidden layer for x_3) + 2x10 (output layer for x_2, x_3).

6. If X is a random variable with a mean of μ and variance σ^2, what is the variance of the random variable $Y = aX + b$, where a and b are constants?

   (a) $a\sigma^2 + b$
   (b) $a^2\sigma^2$
   (c) $a^2\sigma^2 + b^2$
   (d) σ^2

   B. $a^2\sigma^2$

7. Which of the following scenarios is LEAST likely to be modeled accurately by a Gaussian distribution?

   (a) Heights of adult individuals in a large population.
   (b) Scores on a well-designed math test.
   (c) Income distribution in a highly skewed economy.
   (d) Errors in measurements made by a precise sensor.

   C. Income distribution in a highly skewed economy will not be symmetric around the mean. This should follow an exponential distribution.

8. Let X be a continuous random variable with a probability density function $f(x)$. What is the expected value of a function $g(X)$?

   (a) $\int g(x)dx$ over the range of X.
   (b) $\int g(x)f(x)dx$ over the range of X.
   (c) $g(\int xf(x)dx)$ over the range of X.
   (d) $\int xg(f(x))dx$ over the range of X.

   B. $\int g(x)f(x)dx$ over the range of X.

9. If the covariance between two random variables X and Y is zero, what can we conclude?

   (a) X and Y are independent.
   (b) X and Y have no linear relationship.
   (c) The variance of X and Y is zero.
   (d) The mean of X and Y is zero.

   B. X and Y have no linear relationship.

10. A medical test for a disease has a 95% probability of giving a positive result when the person actually has the disease and a 5% probability of giving a positive result when the person does not have the disease. If 1% of the population actually has the disease, what is the probability that a person has the disease given that they have tested positive?

    (a) Approximately 16%
    (b) Approximately 32%
    (c) Approximately 48%
    (d) Approximately 64%

    A. Approximately 16%. $P(D \mid P) = \frac{P(P|D)P(D)}{P(P|D)P(D)+P(P|\bar{D})P(\bar{D})} = \frac{0.95\times0.01}{0.95\times0.01+0.05\times0.99} \approx 0.16$.