CS 261: Deep Generative Models
Quiz 2 Solutions

Available: 01/29/2024; **Due Date:** 23:59 PM PST, 02/02/2024

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 all the content covered last week in the Monday (01/22) and Wednesday (01/24) lectures.
- 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, Feb 02 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. Which of the following statements about Kullback-Leibler (KL) divergence is true?

   (a) KL divergence is always symmetric
   (b) KL divergence is only defined for discrete probability distributions
   (c) $D_{KL}(p, q)$ is undefined when the support of p is strictly greater than the support of q
   (d) $D_{KL}(p, q)$ is undefined when the support of q is strictly greater than the support of p

   C. If the support of p is strictly greater than the support of q, then there exists x for which $p(x) > 0$ and $q(x) = 0$, which will lead to an undefined expression in the log.

2. When does the KL divergence between two distributions become zero?

   (a) When the two probability distributions are identical
   (b) When the two probability distributions have disjoint supports
   (c) When the entropy of the first distribution is zero
   (d) When the entropy of the second distribution is zero

   A. By mathematical definition

3. The expected log-likelihood of a model p_θ under a data distribution p_{data} is equal to the KL divergence between p_{data} and p_θ.

   (a) True
   (b) False

   B. $D_{KL}(p, q) = E_{p_{\text{data}}}[\log p_{\text{data}}] - E_{p_{\text{data}}}[\log p_\theta]$

4. Given an input $\mathbf{x}$, the conditionals in GPT can be evaluated in parallel for density estimation.

(a) True

(b) False

A. True, attention can be evaluated in parallel for all queries and hence the conditionals in GPT can be evaluated in parallel for density estimation.

5. Given an input $\mathbf{x}$, the conditionals in RNN can be evaluated in parallel for density estimation.

   (a) True

   (b) False

   B. False, the hidden state for current timestep depends on the hidden state for the previous timestep creating a sequential dependency.

6. For generating a new sample $\mathbf{x}$, the conditionals in GPT can be sampled in parallel.

   (a) True

   (b) False

   B. False, sampling a conditional in autoregressive models depend on the previous sampled token creating a sequential dependency.

7. For generating a new sample $\mathbf{x}$, the conditionals in RNN can be sampled in parallel.

   (a) True

   (b) False

   B. False, sampling a conditional in autoregressive models depend on the previous sampled token creating a sequential dependency.

8. For two arbitrary random vectors $\mathbf{x}, \mathbf{z}$, there does not exist a joint distribution p such that $p(\mathbf{x}) < p(\mathbf{x}, \mathbf{z})$.

   (a) True

   (b) False

   A. True, $p(\mathbf{x}) = \int_{\mathbf{z}} p(\mathbf{x}, \mathbf{z})$.

9. For any arbitrary random vector $\mathbf{x}$, we have $E_{\mathbf{x}\sim p}[f(\mathbf{x})] = E_{\mathbf{x}\sim q}\left[\frac{f(\mathbf{x})}{q(\mathbf{x})}\right]$.

   (a) True

   (b) False

   B. False, by definition of expectation, $E_{\mathbf{x}\sim p} = E_{\mathbf{x}\sim q}\left[\frac{p(\mathbf{x})f(\mathbf{x})}{q(\mathbf{x})}\right]$.

10. If an autoregressive model assigns high likelihoods to a training set of images, it will also necessarily assign high likelihoods to unseen examples from the same distribution.

    (a) True

    (b) False

    B. False, the model could have overfitted to the training set.