COMP9444
Neural Networks and Deep Learning
Term 2, 2023

Week 3 Tutorial: Probability and Backprop

This page was last updated: 05/27/2023 16:33:59

1. **Bayes' Rule**

   One bag contains 2 red balls and 3 white balls. Another bag contains 3 red balls and 2 green balls. One of these bags is chosen at random, and two balls are drawn randomly from that bag, without replacement. Both of the balls turn out to be red. What is the probability that the first bag is the one that was chosen?

2. **Entropy and KL-Divergence for Discrete Distributions**

   Consider these two probability distributions on the same space $\Omega = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$

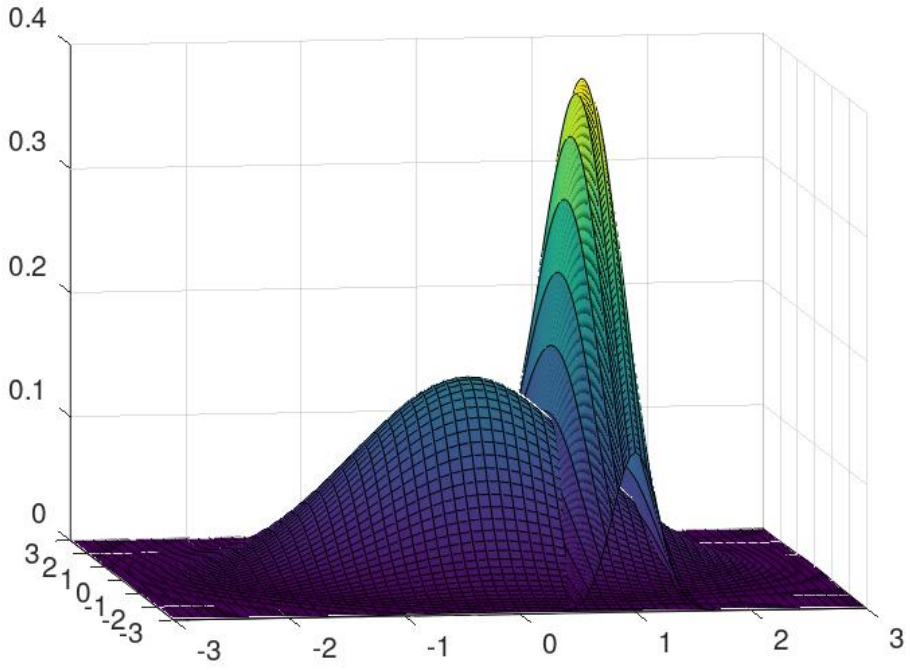
   $p = \langle\ ½, ¼, ⅛, ⅛\ \rangle$
   $q = \langle\ ¼, ⅛, ⅛, ½\ \rangle$

   a. Construct a Huffmann tree for each distribution p and q

   b. Compute the entropy $\mathrm{H}(p)$

   c. Compute the KL-Divergence in each direction $\mathrm{D}_{\mathrm{KL}}(q \parallel p)$ and $\mathrm{D}_{\mathrm{KL}}(p \parallel q)$. Which one is larger? Why?

3. **Entropy, KL-Divergence and W_2 Distance for Bivariate Gaussians**



Consider two bivariate Gaussian distributions p and q.

q has mean $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance $\Sigma_1 = \begin{bmatrix} 0.04 & 0 \\ 0 & 4 \end{bmatrix}$

p has mean $\mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance $\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

a. Compute the Entropy $\mathrm{H}(p)$ and $\mathrm{H}(q)$. Which one is larger? Why?

b. Compute the KL-Divergence in each direction $\mathrm{D}_{\mathrm{KL}}(q \,\|\, p)$ and $\mathrm{D}_{\mathrm{KL}}(p \,\|\, q)$. Which one is larger? Why?

c. Compute the Wasserstein Distance $\mathrm{W}_2(q, p)$

4. **Simple Gradient Descent by Hand**

   Consider the simplest possible machine learning task:

   $$\text{Solve } f(x) = wx \text{ such that } f(1) = 1, \text{ i.e. } f(x) = t, \text{ for } x = 1, t = 1.$$

   We can solve this by gradient descent using the loss function $\mathrm{E} = \tfrac{1}{2}(wx - \mathrm{t})^2$, with learning rate $\eta = 0.5$ and initial value $\mathrm{w} = 0$.

   a. Perform the first epoch of training by completing this table:

      $w = 0$
      $x = 1$
      $f(x) = ?$
      $\mathrm{E} = ?$
      $\partial \mathrm{E}/\partial w = ?$
      $w \leftarrow w - ? = ?$

   b. Repeat these calculations for the second epoch.

   c. Use the code provided in the Exercise in Lesson 2b on Ed to explore what happens for different values of the learning rate. Be prepared to discuss your findings in class.

5. **Any Other Questions**

   Any further questions or discussion about PyTorch, other parts of the course, or broader implications of deep learning.