EE6380/AI2100/AI5100 Deep Learning, Fall 2024

Indian Institute of Technology Hyderabad
Homework 1, Mathematical Preliminaries
50 points. Assigned 21.08.2024, Due **11:59 pm on 31.08.2024**

Everything should be made as simple as possible, but not simpler. – Albert Einstein (?)

Instructions:

- It is **strongly recommended** that you work on your homework on an *individual* basis. If you have any questions or concerns, feel free to talk to the instructor or the TAs.

- Use `matplotlib` to read and plot images - `https://matplotlib.org/tutorials/introductory/images.html`.

- Use `numpy` for basic functions like `log, sqrt, power`. Do not use other built-in functions.

- Download images from University of Southern California's image database at `http://sipi.usc.edu/database/database.php?volume=misc`.

- Please turn in Python Notebooks with the following notation for the file name: `your-roll-number-hw1.ipynb`.

- Do not turn in images. Please use the same names for images in your code as in the database (and as mentioned in the problem statement below). The TAs will use these images to test your code.

Problem Set:

1. **Unit norm ball:** Recall the definition of a unit norm ball from class that we defined in a normed linear space $(X, ||.||)$: $\bar{\mathcal{B}} = \{x \in X : ||x|| \leq 1\}$. Assume $X = \mathbb{R}^2$.

   (a) Write a function that accepts p and plots the $||.||_p$ unit norm ball. Test with integer $p \geq 1$ as well as $0 < p < 1$. (3)

   (b) We claimed that for integer $p \geq 1$ the unit norm ball is convex. Is this clear from the unit norm ball plots? What happens when $0 < p < 1$? Print your observations. (2)

2. **Completeness:** Recall from class that a metric space (X, d) is said to be complete if all Cauchy sequences in X converge to a point in X. Show with a numerical example that the space of continuous functions defined on the closed interval $[0, 1]$ and denoted $C[0, 1]$ is incomplete with respect to the metric derived from the $L1$ norm (i.e., $d(f, g) = ||f - g||_1 = \int_0^1 |f(x) - g(x)| dx$ for any $f, g \in C[0, 1]$).

   Code your example and demonstrate the result either using a plot or numerically. (5)

3. **Entropy of a discrete RV:** Recall the definition of entropy of a discrete RV X from class, $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$, where $p(x)$ is the probability mass function (PMF) of X, and $\mathcal{X}$ is the set of possible values that the random variable X can take.

   (a) Write a function that accepts a PMF as input and outputs the entropy in bits. Do check for the condition of a value being assigned zero probability. (3)

   (b) Now use the above function to plot the entropy of $X \sim \text{Bern}(p)$ as a function of p. Where does this plot attain its maximum? (2)

4. **Joint entropy:** The *joint entropy* of a pair of discrete random variables X and Y with joint distribution $p(x, y)$ is defined as $H(X, Y) = -E_{(x,y) \sim p(x,y)} \log p(X, Y)$ and the *conditional entropy* is defined as $H(Y|X) = -E_{(x,y) \sim p(x,y)} \log p(Y|X)$. Show that: $H(X, Y) = H(X) + H(Y|X)$. (2)

5. **Mutual information:** The *mutual information* between two random variables X and Y is defined as $I(X; Y) = E_{(x,y) \sim p(x,y)} \log \frac{p(X,Y)}{p(x)p(y)}$. Show that $I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$. (3)

6. **Numerical exercise:** If $p(x = 0, y = 0) = p(x = 0, y = 1) = p(x = 1, y = 1) = 1/3, p(x = 1, y = 0) = 0$ (where X, Y are random variables taking values 0 or 1, find

    (a) $H(X), H(Y)$ (2)

    (b) $H(X|Y), H(Y|X)$ (2)

    (c) $H(X, Y)$ (2)

    (d) $H(Y) - H(Y|X)$ (2)

    (e) $I(X; Y)$ (2)

7. **Image entropy:** Download a gray scale image from the link provided in the instructions. By gray scale is meant that the image has one intensity channel. Further, the pixel intensities are in the range $[0, 255]$.

    (a) Write a function that accepts an image as input and returns its normalized histogram. Note that the normalized histogram is found by dividing the original histogram by the total number of pixels in the image. (3)

    (b) Use your entropy function from the earlier problem to find the image entropy. Experiment with different gray scale images from the aforementioned link and note your observations. (2)

8. **Joint PMF and joint entropy:** For this problem, work with the given stereo image pair labeled `left.png` and `right.png` respectively. As in the previous question, both images are gray scale.

    (a) Write a function that accepts this stereo image pair as input, and outputs the normalized joint histogram in addition to plotting it. (3)

    (b) Write a function that accepts the joint PMF of a pair of random variables as input and outputs the joint entropy. (1)

    (c) Test your joint entropy function using the normalized joint histogram computed in Problem 5 (a). (1)

9. **Conditional PMF and conditional entropy:** Continue to work with the stereo image pair.

    (a) Write a function that accepts as input the joint PMF of a pair of random variables, the index of the conditioning random variable, and the value of the conditioning random variable. The function must output the appropriate conditional PMF. (3)

    (b) Write a function that accepts as input the joint PMF and the index of the conditioning RV, and outputs the conditional entropy. (1)

    (c) Test your conditional entropy function using the normalized joint histogram computed in Problem 4 (a), for your choice of the conditioning RV. (1)

10. **KL divergence:** We showed in class that for PMFs p and q defined on $\mathcal{X}$, $D(p||q) \geq 0$, $D(q||p) \neq D(p||q)$. This problem explores these properties experimentally.

    (a) Write a function that accepts as input two PMFs p and q as input, and outputs $D(p||q)$. (1)

    (b) As discussed in class, let $p \sim \text{Bern}(r)$ and $q \sim \text{Bern}(s)$. For a fixed value of r, vary s and do the following: (4)

        i. Plot $D(p||q), D(q||p)$.

        ii. Verify that $D(p||q), D(q||p)$ are indeed non-negative.

        iii. Verify that $D(p||q) \neq D(q||p)$ and are both equal to zero only when $r = s$.

        iv. Finally, find $D(p||q)$ and $D(q||p)$ where p and q are the normalized histograms of `left.png` and `right.png` respectively. Do you think $D(p||q)$ is a good metric for image similarity? Print your response.