Exercise: Information Divergence

(4/4 points)

We now look at a different way to think of Shannon entropy for a random variable X with alphabet $\mathcal{X}$.

Let random variable U have what's called a *uniform distribution* over alphabet $\mathcal{X}$, meaning that

$$p_U(x) = \frac{1}{|\mathcal{X}|} \qquad \text{for all } x \in \mathcal{X}.$$

Notationally, we can write $U \sim \text{Uniform}(\mathcal{X})$.

In the following problems, suppose the number of labels in $\mathcal{X}$ is given by k, i.e., $k = |\mathcal{X}|$.

- What is $H(U)$ in terms of k?

  In this part, please provide your answer as a mathematical formula (and not as Python code). Use `^` for exponentiation, e.g., `x^2` denotes x^2. Explicitly include multiplication using `*`, e.g. `x*y` is xy.

  You may use the function `log`, which for just this part you can treat as log base 2 even though we aren't explicitly writing out the base 2 part. (So for example, `log(x^2)` would be log base 2 of x^2.)

log(k)

Answer: log(k)

Next, we examine the divergence between p_X and the uniform distribution. Show that $D(p_X \parallel p_U)$ can be written of the form

$$D(p_X \parallel p_U) = f(k) - H(X),$$

for a function f that you will determine:

- What is f? Please write your answer in all lowercase with no spaces, and use “log" to mean log base 2 (please do not try to write a subscript 2). Note that we're just asking for what f is, so if your answer is, for instance, `exp`, then just put `exp` and not `exp(k)`.

log

Answer: log

Your answers to the previous two parts should tell you how the entropy of a uniform distribution (over an alphabet of size k) relates to the entropy of any distribution p_X (over the same alphabet of size k).

- Fill in the blanks:

Because of Gibbs' inequality, the entropy of random variable

X

Answer: X

cannot be larger than the entropy of random variable

U

Answer: U

Solution:

- What is $H(U)$ in terms of k?

  Solution:

$$\begin{aligned} H(U) &= \sum_{x \in \mathcal{X}} p_U(x) \log_2 \frac{1}{p_U(x)} \\ &= \sum_{x \in \mathcal{X}} \frac{1}{k} \log_2 \frac{1}{\frac{1}{k}} \\ &= \sum_{x \in \mathcal{X}} \frac{1}{k} \log_2 k \\ &= (\log_2 k) \left(\frac{1}{k}\right) \underbrace{\sum_{x \in \mathcal{X}} 1}_{k} \\ &= \log_2 k. \end{aligned}$$

  So using “log" to mean log base 2, the answer is **log(k)**.

Next, we examine the divergence between p_X and the uniform distribution. Show that $D(p_X \parallel p_U)$ can be written of the form

$$D(p_X \parallel p_U) = f(k) - H(X),$$

for a function f that you will determine:

- What is f? Please write your answer in all lowercase with no spaces, and use "log" to mean log base 2 (please do not try to write a subscript 2). Note that we're just asking for what f is, so if your answer is, for instance, `exp`, then just put `exp` and not `exp(k)`.

  Solution:

$$\begin{aligned} D(p_X \parallel p_U) &= \sum_{x \in \mathcal{X}} p_X(x) \log_2 \frac{p_X(x)}{p_U(x)} \\ &= \sum_{x \in \mathcal{X}} p_X(x) \log_2 \frac{p_X(x)}{1/k} \\ &= \sum_{x \in \mathcal{X}} p_X(x) \log_2 (k p_X(x)) \\ &= \sum_{x \in \mathcal{X}} p_X(x) \log_2 k + \sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x) \\ &= (\log_2 k) \underbrace{\sum_{x \in \mathcal{X}} p_X(x)}_{1} + \sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x) \\ &= \log_2 k - \sum_{x \in \mathcal{X}} p_X(x) \log_2 \frac{1}{p_X(x)} \\ &= \underbrace{\log_2 k}_{f(k)} - H(X). \end{aligned}$$

In particular, f is log base 2, which for this part you answer by just saying **log** to mean log base 2.

Your answers to the previous two parts should tell you how the entropy of a uniform distribution (over an alphabet of size k) relates to the entropy of any distribution p_X (over the same alphabet of size k).

- **Solution:** Because of Gibbs' inequality, the entropy of random variable $\boxed{X}$ cannot be larger than the entropy of random variable $\boxed{U}$.

  In particular, notice that from the answers to the previous parts,

  $$D(p_X \parallel p_U) = H(U) - H(X).$$

  By Gibbs' inequality, information divergence is always nonnegative, which means that we must have $H(U) - H(X) \geq 0$, which means that $H(X) \leq H(U)$.

You have used 2 of 5 submissions