Homework #2

RELEASE DATE: 11/05/2019

DUE DATE: 12/10/2019, BEFORE 17:00 on GRADESCOPE

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FACEBOOK FORUM.

Unless granted by the instructor in advance, you must upload your solution to Gradescope as instructed by the TA.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.

Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.

You should write your solutions in English or Chinese with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.

This homework set comes with 200 points and 20 bonus points. In general, every homework set would come with a full credit of 200 points, with some possible bonus points.

1. Consider the optimal representation of an ensemble X with K alternatives using Huffman coding. Prove that no codeword length n_k is greater than $K-1$ bits.

2. Following the previous question, prove that any string of 1's and 0's is a legal codeword (or a prefix of a legal codeword) of some element $\mathbf{x} \in X^+$.

3. Consider a joint ensemble XY. Prove that $H(X|Y) = H(Y|X)$ if and only if $H(X) = H(Y)$.

4. Consider a "floor" deterministic quantization scheme q on $X = \{0, 1, 2, 3, \ldots, 1126\}$ and $Y = \{0, 1, 2, 3, \ldots, 1126\}$ such that $q(y|x) = 1$ if and only if $y = 2\lfloor x/2 \rfloor$. Prove that for any ensemble X, $H(Y) \le H(X)$.

5. Consider a "round" probabilistic quantization scheme q on $X = \{0, 1, 2, 3, \ldots, 1126\}$ and $Y = \{0, 1, 2, 3, \ldots, 1126\}$ such that

- $q(y = x|x) = 1$ if x is even
- $q(y = x+1|x) = q(y = x-1|x) = 1/2$ if x is odd.

Disprove that for any ensemble X, $H(Y) \le H(X)$ (by constructing a counter example).

6. Consider a joint ensemble XY with probability distribution $p(x, y)$ and a product ensemble $(XY)^2$ with probability distribution $p_2((x_1, y_1), (x_2, y_2)) = p(x_1, y_1)p(x_2, y_2)$ on $((x_1, y_1), (x_2, y_2)) \in (XY)^2$. The product ensemble can be re-arranged as an equivalent product ensemble X^2Y^2 with probability distribution $p_3((x_1, x_2), (y_1, y_2)) = p_2((x_1, y_1), (x_2, y_2))$ for all $((x_1, x_2), (y_1, y_2)) \in X^2Y^2$. Prove or disprove that $I(X^2; Y^2)$ (using p_3) equals $2I(X; Y)$ (using p).

7. Consider an ensemble $X = \{0, 1, 2\}$, where $p(0) = p(1) = p(2) = \frac{1}{3}$. Now approximate X with a set $Y = \{0, 1, 2\}$ with the distortion function $d(x, y) = 0$ except for $d(0, 2) = d(1, 1) = d(2, 0) = 1$.

Taking advantage of the symmetry of the above distortion function, write the general form for $q(y|x)$ that achieves the rate-distortion function, in terms a single parameter $\bar{d} = \epsilon$.

Clarify: We intend to ask you to obtain

$$\text{argmin}_{q:\, \bar{d}_q = \epsilon} I(X; Y_q)$$

where the ensemble Y_q is defined based on the marginal distribution $q(y) = \sum_x p(x)q(y|x)$, and the optimal q is parameterized by ϵ. (*Hint: You can simply mimic page 46 of the IAC notes.*)

8. Let q be the optimal scalar quantization scheme obtained in the previous question. Plot $H(Y_q|X)$, $H(Y_q)$ and $I(X;Y_q)$ with respect to ϵ in the same figure. When does $R(\epsilon) = \min_{q:\,\bar{d}_q \le \epsilon} I(X;Y_q)$ equal $I(X;Y_q)$ in this figure?

9. Consider an ensemble $X = \{0,1\}$ with $0 < p_0 < 1$ denoting the probability of element 0 and $p_1 = 1 - p_0$ denoting the probability of element 1. When proving Shannon's first theorem, we introduced the typical set where

$$T(N,\epsilon) = \left\{ \mathbf{x}\colon \left| \frac{1}{N} \log \frac{1}{p(\mathbf{x})} - H(X) \right| < \epsilon \right\}.$$

Another notion of typicality is to require that $\mathbf{x}$ contains roughly Np_0 0's and Np_1 1's. That is, let $z(\mathbf{x})$ counts the number of 0's in $\mathbf{x}$

$$T_s(N,\delta) = \left\{ \mathbf{x}\colon \left| \frac{1}{N} z(\mathbf{x}) - p_0 \right| < \delta \right\}.$$

Prove that for any given N and δ, there is a corresponding ϵ_δ such that $T_s(N,\delta) \subseteq T(N,\epsilon_\delta)$ and $\lim_{\delta \to 0} \epsilon_\delta = 0$.

T_s is generally called a "strongly typical" set and our earlier T is called a "weakly typical" set. A more general notion of "strongly jointly typical" set can be used to greatly simplify the proof of rate-distortion theorem (and its reverse direction).

10. Following the previous question, if $p_0 \neq p_1$, prove or disprove that for any given N and ϵ, there is a corresponding δ_ϵ such that $T(N,\epsilon) \subseteq T_s(N,\delta_\epsilon)$ and $\lim_{\epsilon \to 0} \delta_\epsilon = 0$.

Note that if $p_0 = p_1 = \frac{1}{2}$, $T(N,\epsilon)$ contains all length-N binary strings while $T_s(N,\delta)$ contains fewer strings when δ is small enough. That is, no such δ_ϵ can exist when $p_0 = p_1$.

Bonus: Conditional Entropy and Prediction Error

11. (Bonus 20 points) In class (actually in the notes), we showed that for a joint ensemble XY, $H(X|Y) = 0$ only if X is a function of Y. That is, $p(x|y) = 1$ or 0 everywhere. Assume that the set X = the set Y, and think of Y as a "guess" of X with $H(X|Y) > \Delta$ for some $\Delta > 0$. That must mean we have some "residual uncertainty" about X after knowing Y. Let $d(x,y) = 1$ if $x \neq y$ and 0 if $x = y$. For any given finite X and any given $\Delta > 0$, prove that there exists a corresponding $\Delta' > 0$ such that for every possible $p(x,y)$ satisfying $H(X|Y) > \Delta$, the average distortion $\sum_{(x,y)} p(x,y)d(x,y) > \Delta'$. (*Hint: d can also be viewed as a random variable.*)