CS 1511 Homework 3

Mathew Varughese, Justin Kramer

Wednesday, Jan 23

5. (a)

When encoding with techniques that involve creating prefix free codes, the letter with higher probabilities of appearing will have shorter bit lengths. The bitlength of a character is $\log_2 1/P(X)$. In this case, $P(x) = \frac{1}{2^k}$, so $\log_2 1/P(X) = \log_2 2^k = k$. In expectation, a letter with a probability of $1/2^k$ of appearing will appear $1/2^k * n$ times in a string of length n where letters are picked from the probability distribution. Therefore, in expectation, each character will have a bitlength of k and show up $n/2^k$ times. Thus, this means the expected total bitlength of a n long string is $\frac{n*k}{2^k}$, which is algebraically equivalent to $n * H(x)$.

5. (b)

In the same idea as the previous question, the bitlength of a character is $\log_2 1/P(X)$. In this new case, an addition is made to so that our probability distribution for each letter which originally appears as $\frac{k}{2^k}$ is now $\frac{k}{2^k} + 1$. In expectation, an individual letter will now contain $\frac{k+2^k}{2^k}$ bits due to the probability of appearing as $1/2^k$ with a bitlength of k that has been appended to in order to make our unpronounced form into the rounded up form of $\frac{1}{2^k}$. Due to there being n letters in our string with these letters being picked from our probability distribution, we expect the total bitlength of the string to be equivalent to $\frac{n(k+2^k)}{2^k}$. Thus, this total bitlength of $\frac{n(k+2^k)}{2^k}$ will be algebraically equivalent to $n * (H(x) + 1)$.

6. (a)

i. $H(X) = \sum P(x) * \log_2 \frac{1}{P(x)}$

$H(X) = \frac{1}{3} * \log_2 3 + \frac{2}{3} * \log_2 \frac{3}{2} = 0.918$

ii. $P(y = 1) = P(y = 1|x = 0) * P(x = 0) + P(y = 1|x = 1) * P(x = 1) = \frac{17}{30}$

$P(y = 0) = P(y = 0|x = 0) * P(x = 0) + P(y = 0|x = 1) * P(x = 1) = \frac{13}{30}$

iii. $H(Y) = \sum P(y) * \log_2 \frac{1}{P(y)}$

$H(Y) = \frac{13}{30} * \log_2 \frac{30}{13} + \frac{17}{30} * \log_2 \frac{30}{17} = 0.9871$

iv. $H(X|Y) = \sum P(x, y) * \log_2 \frac{1}{P(x|y)}$

Using Bayes Formula:

$(((9/13) * log_2(13/9) + (4/13)(log_2(13/4))) * (13/30)) + 17/30 * ((1/17) * (log_2(17) + 16/17 * log_2(17/16))) = 0.561$

v. $H(Y|X) = \sum P(x) * \sum P(y|x) * \log_2 \frac{1}{P(y|x)}$

$1/3*((9/10)*log_2(10/9)+1/10*log_2(10))+2/3*(2/10*log_2(5)+8/10*log_2(10/8)) = 0.63$

vi. $I(X;Y) = H(X) - H(X|Y) = 0.918 - 0.561 = 0.357$

vii. $I(Y;X) = H(Y) - H(Y|X) = 0.9871 - 0.63 = 0.3571$

viii. In this setting this means that there is the same amount of uncertainty between either what is sent or received. If you know what is sent you are a level of "unsure" of what is received. The same logic applies as the other way around.

6. (b)

i.

$I(X;Y)$ is the mutual information between X and Y. This is the measure of the reduction in uncertainity about x that results from learning the value of y. The formula is $I(X;Y) = H(X) - H(X|Y)$. If $I(X;Y) \geq 0$, then $H(X) - H(X|Y) \geq 0$. $H(X) \geq H(X|Y)$. This must be true. The entropy of a distribution must be greater or equal to the conditional entropy. Knowing another probability distribution cannot lower the amount of understood information.

ii.

$I(X;Y) = I(Y;X)$.

$H(X) - H(X|Y) = H(Y) - H(Y|X)$

$H(X|Y) - H(Y|X) = H(Y) - H(X)$

$\sum P(x,y) * \log_2 \frac{1}{P(x|y)} - \sum P(x,y) * \log_2 \frac{1}{P(y|x)} = H(Y) - H(X)$

$\sum P(x,y) * (\log_2 \frac{1}{P(x|y)} - \log_2 \frac{1}{P(y|x)}) = \sum P(x,y) * \log_2 \frac{P(y|x)}{P(x|y)} = H(Y) - H(X)$

Apply Bayes Theorem. $\frac{P(y|x)}{P(x|y)} = \frac{P(x|y)*(P(y)/P(x)}{P(x|y))} = \frac{P(y)}{P(x)}$

$\sum P(x,y) * \log_2 \frac{P(y)}{P(x)} = H(Y) - H(X)$

$\sum P(x,y) * \log_2 \frac{P(y)}{P(x)} = \sum P(y) * \log_2 \frac{1}{P(y)} - \sum P(x) * \log_2 \frac{1}{P(x)}$

Unsure how to proceed from here, especially since the bounds of the summations are different. Logically it makes sense that the mutual information between X and Y is the same as the mutual information between Y and X. Not sure exactly how to prove that fully mathematically.