ANLY601 Take Home Assignment 3

Xunge Jiang

Feb 24, 2019

1 Assume a discrete random variable X has 4 possible outcomes with corresponding probabilities:

$$p(x_1) = .1$$
$$p(x_2) = .4$$
$$p(x_3) = .2$$
$$p(x_4) = .3$$

1.1 a) What is the entropy of random variable X?

Log base is 2:

$$\begin{aligned} H(X) &= -\sum_X p(X) logp(X) \\ &= -(0.1 \times log_2 0.1 + 0.4 \times log_2 0.4 + 0.2 \times log_2 0.2 + 0.3 \times log_2 0.3) \\ &= 1.85 \ bits \end{aligned}$$

1.2 b) Which of the 4 observations conveys the most information?

The event with low probability carries more information. Thus, the first observation conveys the most information.

1.3 c) Compare the entropy of X to a random variable Z with 4 outcomes that have uniform probability.

For random variable Z, each event has probability $p(x_i) = 0.25$ where i = 1, 2, 3, 4.

$$H(Z) = -\sum_Z p(Z) logp(Z) = -(0.25 \times log_2 0.25) * 4 = 2 \ bits$$

Since $H(Z) > H(X)$, the entropy of Z is higher than the entropy of X.

2 Prove the following relationship between the differential entropy, conditional entropy and entropy holds, using the definitions of $H[y|x]$ and $H[x]$. $H[x, y] = H[y|x] + H[x]$

Proof:

$$\begin{aligned}
H[x,y] &= -\sum_x \sum_y p(x,y) log p(x,y) \\
&= -\sum_x \sum_y p(x,y) log p(x) p(y|x) \\
&= -\sum_x \sum_y p(x,y) log p(x) - \sum_x \sum_y p(x,y) log p(y|x) \\
&= -\sum_x p(x) log p(x) \sum_y p(y|x) - \sum_x \sum_y p(x,y) log p(y|x) \\
&= -\sum_x p(x) log p(x) - \sum_x \sum_y p(x,y) log p(y|x) \\
&= H[x] + H[y|x]
\end{aligned}$$