12.

Intro

In this question, we will demonstrate the relationship between mutual information and entropy. Before start the solution, remember that the mutual information between two r.v's is a more general way of measuring the relationship between them, than the correlation coeficient. On the other hand, entropy is a measure of uncertainty. Thus, we expect two situations. Or the mutual information is zero, and the conditional entropy do not change. Or the mutual information is positive and the conditional entropy is less than the prior entropy. In another words, the mutual information is a measure of how the knowledge of one r.v's takes away the uncertainty of the other r.v.

Solution

We want to show that:

$$I(X,Y) = H(X) - H(X|Y) \tag{1}$$

Let's start by expanding the three terms:

$$I(X,Y) = \sum_x \sum_y p_{xy} log\left(\frac{p_{xy}}{p_x p_y}\right) \tag{2}$$

$$H(X) = -\sum_x p_x log(p_x) \tag{3}$$

$$H(X|Y) = -\sum_y p_y H(X|Y=y) \tag{4}$$

Looking (2), (3) and (4) we can get a nice tip. (2) is an expression involving summations on x and y. On the other hand, (3) and (4) are expressions involving summations on just one of the r.v's. So, we should expand (3) and (4) to a double summation on x and y. To expand (3) we just have to realize that p_x is the marginalization of p_{xy} with respect to y:

$$H(X) = -\sum_x p_x log(p_x) = -\sum_x \sum_y p_{xy} log(p_x) \tag{5}$$

To expand (4), we need to substitute $H(X|Y)$ by its expression:

$$H(X|Y) = \sum_y p_y H(X|Y=y) = -\sum_y \sum_x p_y p_{x|y} log(p_{x|y}) = -\sum_y \sum_x p_{xy} log\left(\frac{p_{xy}}{p_y}\right) \tag{6}$$

The final passage consists of combining (5) and (6) and compare with (2):

$$H(X) - H(X|Y) = -\sum_x \sum_y p_{xy} log(p_x) + \sum_y \sum_x p_{xy} log\left(\frac{p_{xy}}{p_y}\right) = \sum_x \sum_y p_{xy} log\left(\frac{p_{xy}}{p_x p_y}\right) = I(X,Y) \tag{7}$$

I prefer to express this result as:

$$H(X|Y) = H(X) - I(X,Y) \tag{8}$$

Since the mutual information satisfies the property of symmetry $I(X,Y) = I(Y,X)$, the expression $H(Y|X) = H(Y) - I(X,Y)$ is also true.

Conclusion

In this question, we demonstrate the relationship between the mutual information and the entropies of two r.v's. The relationship shows us how to update the entropy of one r.v. given its mutual information with another r.v. We can think on this relationship as a version of Baye's rule for information theory. We start with a prior $H(X)$, get more information with $I(X,Y)$ and update our beliefs $H(X|Y)$. Note that the new information given by the mutual information can be expressed as $I(X,Y) = H(Y) - H(Y|X)$. So the posterior entropy can be expressed as:

$$H(X|Y) = H(X) + H(Y|X) - H(Y) \tag{9}$$

The main difference between (9) and Bayes' rule is that instead of multiply/divide terms, we add/subtract them. This distinction occurs because entropy and mutual information work with the **log** function, turning multiplication into addition.