14.

Intro

In this question, we are going to analyse the meaning of r, a function defined in terms of entropies. As the text hints, we should expect that r behaves like a information theory correlation funcion. In another words, r should be a normalized value that tell how dependent two r.v.'s are from each other.

Solution

a)

We must show that $r = \frac{I(X,Y)}{H(X)}$:

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

b)

$$H(p) = \sum p(-logp) \geq 0 \tag{2}$$

Using (2), we can write:

$$\frac{H(Y|X)}{H(X)} \geq 0 \implies 1 - \frac{H(Y|X)}{H(X)} \leq 1 \tag{3}$$

Also,

$$I(X,Y) = H(X) - H(Y|X) \geq 0 \implies H(Y|X) \leq H(X) \implies 1 - \frac{H(Y|X)}{H(X)} \geq 0 \tag{4}$$

Thus, $0 \leq r \leq 1$

c)

$r = \frac{I(X,Y)}{H(X)} = 0 \iff I(X,Y) = 0$

d)

$r = 1 \iff I(X,Y) = H(X)$. Which is the same that saying $r = 1 \iff H(Y|X) = 0$. Let's see what $H(Y|X) = 0$ means:

$$\begin{aligned} H(Y|X) = \sum_x p(x)H(Y|X = x) = \sum_x p(x) \sum_y p(y|x)(-log(p(y|x)) = \\ \sum_x \sum_y p(x)p(y|x)(-log(p(y|x)) \end{aligned} \tag{5}$$

Observe that each term of the summation is non negative. Thus, every term must be 0 in order to (5) to be true. Thus, whether $p(y|x) = 0$ or $logp(y|x) = 0$. The second condition only happens when $p(y|x) = 1$. So, $H(Y|X) = 0$ say to us that given X, we know exactly which Y is going to occur. In another words, $Y = f(X)$. Since the problem is symmetrical, f is invertible and we can also express this as $X = f^{-1}(X)$

Conclusion

In this section we study the parameter r defined as $r = 1 - \frac{H(Y|X)}{H(X)}$. We discored that it behaves very similarly to the classical correlation. Being more specific, we found out four things. First, r is a ratio between mutual information and entropy, much like the classical correlation is a ration between covariance and variances (Note that since we are assuming $H(X) = H(Y)$ then $r = \frac{I(X,Y)}{\sqrt{H(X)}\sqrt{H(Y)}}$). Second, the ratio is normalized (i.e. between zero and unity). Third, $r = 0 \iff X$ and Y are independet. Fourth, we discored that $r = 1 \iff Y = f(X)$, where f is invertible. Overall, we see that the parameter r has the same goal as ρ, but it is more general, because it uses entropies instead of covariances. The first advantege is that it captures linear and non linear functions between X and Y. The second advantege is that $r = 0 \implies X$ and Y are independent.