TP1

Thibault Duhamel 18026048 et Heng Shi 18171434

February 2, 2019

Question 1

We here chose to introduce discrete variables to demonstrate the following property of the entropy:

$$H[x, y] = H[y|x] + H[x]$$

Let us start from the right side of the equation, which is the most complex, to arrive to the left side:

$$H(y|x) + H(x) = -\sum_{x}\sum_{y} p(x, y)log(p(y|x)) - \sum_{x} p(x)log(p(x))$$

The sum is a linear operation, which leads us to merge the two parts:

$$= -\sum_{x}\Big[\sum_{y} p(x, y)log(p(y|x)) + p(x)log(p(x))\Big]$$

It is useful to introduce the equality of conditional probability:

$$= -\sum_{x}\Big[\sum_{y} p(x, y)log(\frac{p(x, y)}{p(x)}) + p(x)log(p(x))\Big]$$

We are now able to break down the logarithm term as follows, by the use of its properties:

$$= -\sum_{x}\Big[\sum_{y} p(x, y)[log(p(x, y)) - log(p(x))] + p(x)log(p(x))\Big]$$

Then, we do expand the items according to distributive property:

$$= -\sum_{x}\Big[\sum_{y} p(x, y)log(p(x, y)) - \sum_{y} p(x, y)log(p(x)) + p(x)log(p(x))\Big]$$

Moreover, the second factor can be added based on the variable y, which is the property of joint probabilities:

$$= - \sum_x \Big[\sum_y p(x,y) log(p(x,y)) - p(x) log(p(x)) + p(x) log(p(x)) \Big]$$

$$= - \sum_x \sum_y p(x,y) log(p(x,y))$$

What happens to be exactly the definition of the joined entropy:

$$\boxed{H(y|x) + H(x) = H(x,y)}$$

Question 2

We here also decided to introduce discrete variables to demonstrate the following property of mutual information:

$$I(x,y) = H(x) - H(x|y)$$

Let us start from the right side of the equation to get to the left side:

$$H(x) - H(x|y) = - \sum_x p(x) log(p(x)) + \sum_x \sum_y p(x,y) log(p(x|y))$$

Using the linearity of the sum, we will transform it into the following equation:

$$= \sum_x \Big[- p(x) log(p(x)) + \sum_y p(x,y) log(p(x|y)) \Big]$$

We can now replace the conditional probability by its relation to the joint probability:

$$= \sum_x \Big[- p(x) log(p(x)) + \sum_y p(x,y) log(\frac{p(x,y)}{p(y)}) \Big]$$

The first term of the sum can be marginalized over the variable y:

$$= \sum_x \Big[- \sum_y p(x,y) log(p(x)) + \sum_y p(x,y) log(\frac{p(x,y)}{p(y)}) \Big]$$

Now, the two sums of variable y can be merged:

$$= \sum_x \sum_y \Big[- p(x,y) log(p(x)) + p(x,y) log(\frac{p(x,y)}{p(y)}) \Big]$$

Finally, the property of logarithms allows us to write it as follows:

$$= \sum_x \sum_y \left[p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \right]$$

Which is exactly the definition of the mutual information and therefore proves the property:

$$\boxed{H(x) - H(x|y) = I(x,y)}$$

Question 3

In the question, we also demonstrate the following property in the context of discrete variables:

$$Cov[x,y] = E_{xy}[xy] - E[x]E[y]$$

The covariance is defined by the following equation:

$$Cov[x,y] = E[(x - E[x])(y - E[y])]$$

Let us distribute each item of the product:

$$= E[xy - xE[y] - yE[x] + E[x]E[y]]$$

The linearity of the expectation allows us to split it into four parts:

$$= E[xy] - E[xE[y]] - E[yE[x]] + E[E[x]E[y]]$$

Since E[x] and E[y] are constants, the linearity can herein be used again:

$$= E[xy] - E[y]E[x] - E[x]E[y] + E[x]E[y]$$

$$= E[xy] - E[y]E[x]$$

As a conclusion, the covariance definition is linked by the following equalities:

$$\boxed{Cov[x,y] = E[xy] - E[y]E[x] = E_{xy}[x,y] - E[x]E[y]}$$