Lecture 3

Appendix 1
Computation of the conditional entropy

Example

Let (X,Y) have the following joint distribution

Y \ X	1	2
1	0	3/4
2	1/8	1/8

Then $H(X)=(1/8, 7/8)=0,544$ bits, $H(X|Y=1)=0$ bits and $H(X|Y=2)=1$ bit. We calculate $H(X|Y)=3/4\ H(X|Y=1)+1/4\ H(X|Y=2)=0.25$ bits. Thus the uncertainty in X is increased if Y=2 is observed and decreased if Y=1 is observed, but uncertainty decreases on the average.

Computation of Conditional Entropy

To compute the conditional entropy in this example, we recall its definition:

$$H(X \mid y = b) = \sum_x p(x \mid y = b) \log \frac{1}{p(x \mid y = b)}$$

Where p(x|y=b) is the conditional probability, defined as:

$$p(x = a \mid y = b) = \frac{p(x = a, y = b)}{p(y = b)}$$

Where p(x=a,y=b) is the joint probability, that is represented in a matrix such as the one in the previous example, and where p(y=b) is the marginal probability:

$$p(y = b) = \sum_x p(x, y = b)$$

That is, the probability of X taking on a particular value x=a is the sum of the joint probabilities of this outcome for X and all possible outcomes for Y (summing the elements in the corresponding row of the matrix in the example).

Conditional Entropy (Example)

Referring to the example in the first slide, we have that:

$$H(X \mid y = 1) = \sum_x p(x \mid y = 1) \log \frac{1}{p(x \mid y = 1)}$$

$$= p(x = 1 \mid y = 1) \log \frac{1}{p(x = 1 \mid y = 1)} + p(x = 2 \mid y = 1) \log \frac{1}{p(x = 2 \mid y = 1)}$$

$$= \frac{p(x = 1, y = 1)}{p(y = 1)} \log \frac{p(y = 1)}{p(x = 1, y = 1)} + \frac{p(x = 2, y = 1)}{p(y = 1)} \log \frac{p(y = 1)}{p(x = 2, y = 1)}$$

$$= \frac{0}{3/4} \log \frac{3/4}{0} + \frac{3/4}{3/4} \log \frac{3/4}{3/4} = 0$$