INFORMATION AND CODING THEORY

ELEN060-2

Destexhe Clara S191103
Hogge Louis S192814

Project 1 - Information measures

March 15, 2023

1 Implementation

Question 1 : Entropy

Entropy is a measure of uncertainty in a random variable. It quantifies the amount of information a random variable is expected to contain. It is computed in bits due to the log in base two. The minus is added to counter the logarithm value (always negative). Indeed, entropy is never negative.

If $\mathcal{X}$ is a discrete random variable and $P(X_i)$ the value of the probability distribution of the event $\mathcal{X}_i$, then the entropy of $\mathcal{X}$ is :

$$H(\mathcal{X}) \triangleq -\sum_{i=1}^{n} P(\mathcal{X}_i) \log_2 P(\mathcal{X}_i)$$

In order to compute the entropy, a sum is done on the elements of the probability distribution vector. Zero probabilities are not considered in the sum, thanks to the added condition in the loop.

Question 2 : Joint Entropy

The joint entropy measures of uncertainty related to a set of random variables. The entropy and the joint entropy have the same formula but the joint probability matrix replaces the probability vector. The matrix is a 2-D array; thus, a sum is added to consider each element in the formula.

If $\mathcal{X}$ and $\mathcal{Y}$ are discrete random variables, $P(\mathcal{X}_i \cap \mathcal{Y}_j)$ is the probability of the random variable $\mathcal{X}$ to be $\mathcal{X}_i$ and in the same time the random variable $\mathcal{Y}$ to be $\mathcal{Y}_j$. The joint entropy is then defined as :

$$H(\mathcal{X}, \mathcal{Y}) \triangleq -\sum_{i=1}^{n}\sum_{j=1}^{m} P(\mathcal{X}_i \cap \mathcal{Y}_j) \log_2 P(\mathcal{X}_i \cap \mathcal{Y}_j)$$

In order to compute the joint entropy with the probability distribution $P(\mathcal{X}_i \cap \mathcal{Y}_j)$, a double loop is implemented to pass on each element of the probability matrix. As for the entropy, the zero probabilities are not taken into account.

Question 3 : Conditional Entropy

The conditional entropy measures the uncertainty related to the outcome of an event knowing the value of the other random variable. The formula is the same as the joint entropy but the conditional probability distribution replaces the joint probability distribution.

If $\mathcal{X}$ and $\mathcal{Y}$ are discrete random variables, $P(X_i \mid Y_j)$ is the conditional probability meaning the probability of $\mathcal{X}$ given the outcome of $\mathcal{Y}$. Then, the conditional entropy is defined as :

$$H(\mathcal{X} \mid \mathcal{Y}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X_i \cap Y_j) \log_2 P(X_i \mid Y_j)$$

In order to compute the conditional entropy with the probability distribution $P(X_i \cap Y_j)$, the formula needs to be rewritten as in slide 4 of the second course :

$$H(\mathcal{X} \mid \mathcal{Y}) = H(\mathcal{X}, \mathcal{Y}) - H(\mathcal{Y})$$
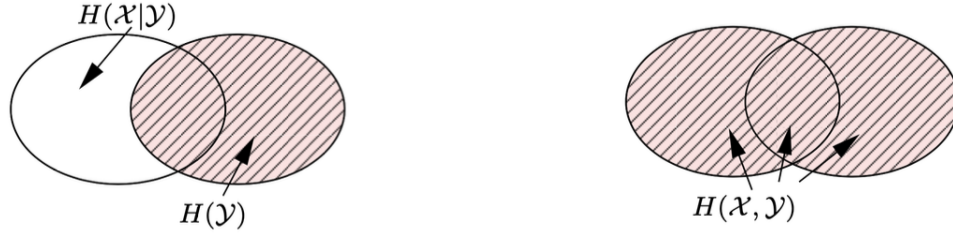
It can be proved graphically in the figure 1 :

Figure 1: Conditional entropy

To implement this, the joint entropy function is applied to the joint probability distribution and the entropy function is applied to the sum on axis zero of the joint probability distribution to obtain $P(\mathcal{Y})$.

Question 4 : Mutual Information

The mutual information measures the dependence between a set of random variables. If the set of random variables is independent, the mutual information is equal to zero. The higher this measure, the higher the dependency between the random variables.

If $\mathcal{X}$ and $\mathcal{Y}$ are discrete random variables, $P(X_i \cap Y_j)$ is the probability of the random variable $\mathcal{X}$ to be $\mathcal{X}_i$ and in the same time the random variable $\mathcal{Y}$ to be $\mathcal{Y}_j$. Also, $P(\mathcal{X}_i)$ is the probability of the event $\mathcal{X}_i$ and $P(\mathcal{Y}_j)$ the probability of the event $\mathcal{Y}_j$. Then, the mutual information is defined as :

$$I(\mathcal{X};\mathcal{Y}) = +\sum_{i=1}^{n}\sum_{j=1}^{m} P(\mathcal{X}_i \cap \mathcal{Y}_j) \log \frac{P(\mathcal{X}_i \cap \mathcal{Y}_j)}{P(\mathcal{X}_i) P(\mathcal{Y}_j)}$$

In order to compute the mutual information with the probability distribution $P(\mathcal{X}_i \cap \mathcal{Y}_j)$, the formula needs to be rewritten as in slide 6 of the second course :

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X},\mathcal{Y})$$

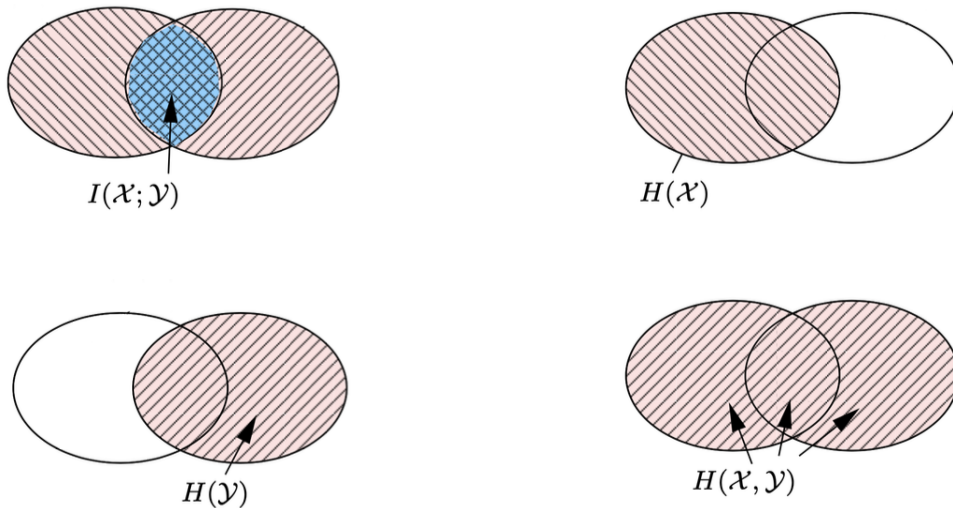It can be proved graphically in the figure 2 :

Figure 2: Mutual Information

To compute all the measures needed in the formula above, $P(\mathcal{X})$ and $P(\mathcal{Y})$ need to be defined. A sum is done on the joint probability matrix, on axis zero to obtain $P(\mathcal{Y})$ and on axis one to obtain

$P(\mathcal{X})$. Then, the entropy function is applied to these two vectors and the joint entropy function is applied to the joint probability distribution.

Question 5 : Conditional joint entropy and conditional mutual information

Conditional joint entropy

If $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are three random variables, the conditional joint entropy of $\mathcal{X}$, $\mathcal{Y}$ given $\mathcal{Z}$ is defined as :

$$H(\mathcal{X}, \mathcal{Y} \mid \mathcal{Z}) = H(\mathcal{X} \mid \mathcal{Z}) + H(\mathcal{Y} \mid \mathcal{X}, \mathcal{Z})$$

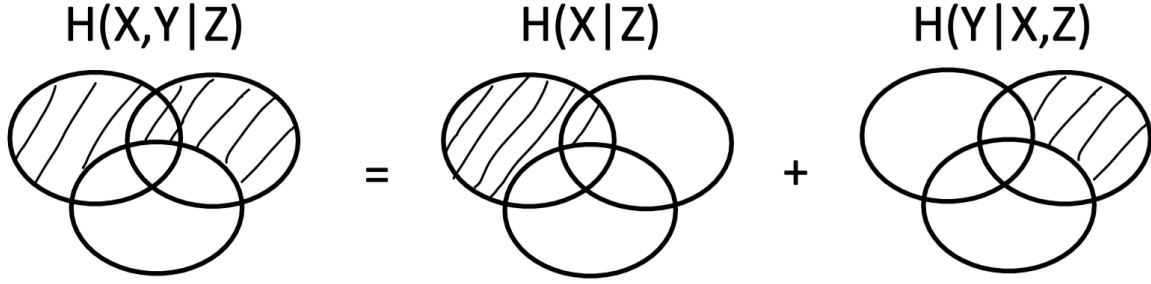It can be proved graphically in the figure 3 :

Figure 3: Conditional joint entropy

To implement this formula, the entropy function is used to compute $H(\mathcal{X} \mid \mathcal{Z})$. The conditional entropy $H(\mathcal{Y} \mid \mathcal{X}, \mathcal{Z})$ is computed thanks to the Bayes theorem :

$$P(Y \mid X, Z) = \frac{P(\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z})}{P(\mathcal{X} \cap \mathcal{Z})}$$

Conditional mutual information

If $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are three random variables, the conditional mutual information of $\mathcal{X}$, $\mathcal{Y}$ given $\mathcal{Z}$ is defined as on slide 10 of the second course :

$$I(\mathcal{X}; \mathcal{Y} \mid \mathcal{Z}) \triangleq H(\mathcal{X} \mid \mathcal{Z}) - H(\mathcal{X} \mid \mathcal{Y}, \mathcal{Z})$$

To implement this formula, to compute $H(\mathcal{X} \mid \mathcal{Z})$, the conditional entropy function is applied on the joint probability distribution of $\mathcal{X}$ and $\mathcal{Z}$. $H(\mathcal{X} \mid \mathcal{Y}, \mathcal{Z})$ needs to be rewritten as :

$$H(\mathcal{X} \mid \mathcal{Y}, \mathcal{Z}) = -\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{l} (P(\mathcal{X}_i \cap \mathcal{Y}_j \cap \mathcal{Z}_k)) \log_2(P(\mathcal{X}_i \mid \mathcal{Y}_j, \mathcal{Z}_k))$$

where $P(\mathcal{X}_i \mid \mathcal{Y}_j, \mathcal{Z}_k)$ can be rewritten thanks to the Bayes rule as :

$$P(\mathcal{X} \mid \mathcal{Y}, \mathcal{Z}) = \frac{P(\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z})}{P(\mathcal{Y} \cap \mathcal{Z})}$$

2 Predicting the outcome of a football game

Question 6 : Entropy and cardinality

In the table 1, the entropy and the cardinality of each variable can be found.

Variable	Cardinality	Entropy
outcome	3	1.33488
previous_outcome	3	1.483
day	7	2.80656
time	3	0.93252
month	12	3.58263
wind_speed	3	1.58471
weather	4	1.76408
location	2	0.99994
capacity	4	1.53391
stadium_state	2	0.63955
injury	2	0.99984
match_type	2	0.9999
opponent_strength	3	1.58435

Table 1: Cardinality and entropy of each variable

Variables with a higher/lower cardinality typically have higher/lower entropy, which corresponds to higher/lower levels of uncertainty or randomness in the values of those variables. For instance, in contrast to the *location* variable, the *month* variable has a much higher entropy value.

If a variable's values are distributed evenly/not evenly for a given cardinality, the entropy may be higher/lower. For instance, with a cardinality of 2, the *stadium_state* variable has a lower entropy than the other variables with the same cardinality because the value *dry* is much more prevalent than the value *wet*.

Theoretically, if we look at the entropy's formula:

$$H(\mathcal{X}) \triangleq -\sum_{i=1}^{n} P(\mathcal{X}_i) \log_2 P(\mathcal{X}_i)$$

If a discrete random variable $\mathcal{X}$ has a larger/lower cardinality, then there are more/fewer potential values of $\mathcal{X}$ and the probabilities of these values are probably more dispersed/concentrated. As a result, the entropy value increases/decreases.

Moreover, the distribution of probabilities affects the entropy value. If the probabilities are evenly distributed, because each value of $\mathcal{X}$ is equally likely and there is more uncertainty, the entropy value will be larger. If they are not evenly distributed, because some values of $\mathcal{X}$ are more likely than others and there is less uncertainty, the entropy value will be lower.

Question 7 : Conditional entropy of *outcome*

As it can be seen in the tables 2 and 3, the conditional entropy of *outcome* is very near to the entropy of *outcome* when *wind_speed* is the conditioning variable. It shows that *wind_speed* has little impact on the uncertainty or randomness of *outcome*. According to this, *wind_speed* is not a reliable predictor of *outcome*.

The conditional entropy of *outcome* is lower when *previous_outcome* is the conditioning variable. It shows that *previous_outcome* is a powerful predictor of *outcome*. Thus, a performance's past behavior is a reliable predictor of its present behavior.

Variable	Conditional entropy
previous_outcome	1.18148
day	1.33349
time	1.3338
month	1.33036
wind_speed	1.33473
weather	1.33384
location	1.33351
capacity	1.33202
stadium_state	1.33432
injury	1.33024
match_type	1.33483
opponent_strength	0.93861

Table 2: Conditional entropy of *outcome* given each of the other variables

Variable	Entropy
outcome	1.33488

Table 3: Entropy of *outcome*

Question 8 : Mutual information between *month* and *capacity*

The table 4 displays the mutual information between *month* and *capacity* as well as the mutual information between *day* and *time*.

Variables	Mutual information
month and capacity	0.00607
day and time	0.50461

Table 4: Mutual information

With a mutual information of only 0.00607, it can be concluded that there is a small correlation between *month* and *capacity*. In other words, little knowledge about the value of one variable does not reveal much about the value of the other. This implies these factors are largely independent of one another, and they might not even need to be taken into account simultaneously when developing predictive models.

On the other hand, the mutual information between *day* and *time* is fairly high at 0.50461, showing a close association between these two variables. A large amount of information regarding the value of the other variable can be learned from knowing the value of one variable. This implies these variables are interconnected and might need to be taken into account together when developing predictive models.

Question 9 : Conditional entropy and mutual information of *outcome*

As can be seen in table 5, according to the mutual information, *opponent_strength* has the highest value, coming in at 0.39627. Hence, if we could only invest in one variable, we would do it on *opponent_strength*.

Also, if the choice is based on conditional entropy, *opponent_strength* is the variable that implies the lowest conditional entropy of *outcome* with 0.93861. As a result, if conditional entropy is used to make the decision, we would also select *opponent_strength*.

Variable	Conditional entropy	Mutual information
previous_outcome	1.18148	0.1534
day	1.33349	0.00139
time	1.3338	0.00108
month	1.33036	0.00452
wind_speed	1.33473	0.00015
weather	1.33384	0.00104
location	1.33351	0.00137
capacity	1.33202	0.00286
stadium_state	1.33432	0.00055
injury	1.33024	0.00464
match_type	1.33483	0.00005
opponent_strength	0.93861	0.39627

Table 5: Conditional entropy and mutual information of *outcome* given each of the other variables

Question 10 : Conditional mutual information of *outcome*

The table 6 displays the mutual information between *outcome* and each other variables given *previous_outcome*.

Variable	Mutual information
day	0.00474
time	0.00347
month	0.01369
wind_speed	0.00212
weather	0.00303
location	0.00213
capacity	0.00495
stadium_state	0.00062
injury	0.009
match_type	0.00051
opponent_strength	0.24459

Table 6: Mutual Information between *outcome* and each other variables given *previous_outcome*

The *opponent_strength* variable is still the best to invest in given the new information that the outcome of previous matches between the same opponent are now freely available. As it can be seen in the new data, while taking into consideration the outcome of previous matches, the mutual information between *outcome* and *opponent_strength* reduces from 0.39627 to 0.24459 which shows that some of the knowledge provided by *opponent_strength* about *outcome* is contained in *previous_outcome*. Moreover the amount of information provided by all the other variables increases when *previous_outcome*'s information is taken into consideration. But 0.24459 is still the highest. The *opponent_strength* variable is therefore still the most useful for forecasting the outcome of the game.

Question 11 : Mutual information of *location*

Variables	Mutual information
stadium_state and weather	0.0

Table 7: Mutual information when *location* is home

As can be seen in the table 7, when the *location* is home, the mutual information between *stadium_state* and *weather* is equal to 0.0, indicating that the two variables are independent of one another.

It's conceivable that the home stadium has a unique microclimate unaffected by the weather outside. For instance, the stadium might be built in a protected place or have a roof over it to keep the weather

out. Despite the fact they might be associated in other places, this thus results in the *stadium_state* variable being independent of the *weather* variable.

3 Playing with information theory-based strategy

Question 12 : Entropy computation

In this situation, all colors have the same probability to be on the board. The probability distribution for one slot is uniform and written as $P(X_i) = \frac{1}{6}$ for i = 1, ..., 6. Each slot is independent. Therefore, their entropy is the same.

$$H(X) = -log_2\left(\frac{1}{6}\right) = 2,59$$

The probability of the whole game for each color is the same. The game can accept 6^4 combinations. The probability distribution for the whole game is uniform and is written as $P(X_i) = \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} = (\frac{1}{6})^4$ with i = 1, ..., 6 because the whole game is made of four slots. The entropy is then :

$$H(X) = -log_2\left(\left(\frac{1}{6}\right)^4\right) = 10,34$$

The two entropies are linked by the factor 4, due to the four slots present in the whole game.

Question 13 : First Guess

Firstly, the slot for which the guess is right, its entropy is equal to zero since the probability for this slot for this color is equal to one and the others are zero. Secondly, two cases are possible. The first case is if the correct color is the red or the yellow slot. If it is the case, each cell has now a probability of $\frac{1}{4}$ for each color. For instance, if the first slot is the correct one, the blue slots and the green slot can be neither green or blue. Four colors are left to complete the slots. The entropy for the wrong slots is then :

$$H(\text{Each slot left}) = -log_2\left(\frac{1}{4}\right) = 2$$

.

And the entropy for the whole board is :

$$H(\text{Whole board}) = 3 * H(\text{Each slot left}) = 6$$

The second case is if the correct color is one of the blue slot. Then, the red, green and the other slot can neither be blue, green or red. The slots have only three colors left to be. Each cell has now a probability of $\frac{1}{3}$ for each color. In this case, the entropy is

$$H(X_i) = -log_2\left(\frac{1}{3}\right) = 1,58$$

And the entropy for the whole board is :

$$H(\text{Whole board}) = 3 * H(\text{Each slot left}) = 4,74$$

Now, the entropy needs to be generalized for all possibilities. The expectation of the two entropies will give the generalized value for the entropy.

$$H(\text{Whole board}) = \frac{6 + 4,74}{2} = 5,37$$

The information obtained is H_{stage0} - H_{stage1} = 10,34 − 5,37 = 4,97, meaning the information can be represented in 5 bits.

Question 14 : First guess bis

For this question, a lot of cases are possible. Some of the cases have the same entropy. First, as before, the slots that have been correctly guessed have always an entropy equal to zero.

First case, the slot one or three are the correct ones and one of the blue slot is the incorrect one. This situation happens four times. The table 8 shows the probability distribution of one of this situation :

Colours / Slot number	1	2	3	4
1	1	1/4	0	1/4
2	0	0	1	1/4
3	0	1/4	0	1/4
4	0	0	0	0
5	0	1/4	0	1/4
6	0	1/4	0	1/4
Entropy	0	2	0	2

Table 8: Slot 1 is correct and slot 2 is incorrect

Thus, in these cases the total entropy is equal to 4.

Second case, the slot one or three are the correct or the incorrect slots. This case happens two times. The table 9 shows the probability distribution of one of this situation.

Colours / Slot number	1	2	3	4
1	1	1/8	1/4	1/8
2	0	0	0	0
3	0	1/8	1/4	1/8
4	0	1/2	0	1/2
5	0	1/8	1/4	1/8
6	0	1/8	1/4	1/8
Entropy	0	2	2	2

Table 9: Slot 1 is correct and slot 3 is incorrect

Thus, in these cases the total entropy is equal to 6.

Third case, one of the blue slot is the correct one and the slots 1 or 3 is the incorrect one. This situation happens four times. The table 10 shows the probability distribution of one of this situation :

Colours / Slot number	1	2	3	4
1	0	0	1/2	1/2
2	0	1	0	0
3	1/3	0	1/6	1/6
4	0	0	0	0
5	1/3	0	1/6	1/6
6	1/3	0	1/6	1/6
Entropy	1,58	0	1,79	1,79

Table 10: Slot 2 is correct and slot 1 is incorrect

Thus, in these cases the total entropy is equal to 5,16.

Last case, the blue slots are the correct and the incorrect one. This situation happens in two cases. The table 11 represents the probability distribution of one of this situation :

Colours / Slot number	1	2	3	4
1	0	0	0	0
2	1/2	1	1/2	0
3	1/6	0	1/6	1/3
4	0	0	0	0
5	1/6	0	1/6	1/3
6	1/6	0	1/6	1/3
Entropy	1,79	0	1,79	1,58

Table 11: Slot 2 is correct and slot 4 is incorrect

Thus, in these cases the total entropy is equal to 5,16.

The whole entropy is the sum of each slot's entropy. All of these cases need to be generalized to one unique entropy. The expectation of these cases will give the generalized entropy :

$$H(\text{Whole board}) = \frac{4*4 + 2*6 + 6*5,16}{12} = 4,91$$

The information obtained is H_{stage0} - $H_{stage1} = 10,34 - 4,91 = 5,43$, meaning the information can be represented in 5 bits.

Question 15 : Maximum entropy

For one slot, with $\mathcal{C}$ as the number of possible colors, the formula would become :

$$H(\mathcal{X}_i) = -\log_2\left(\frac{1}{\mathcal{C}}\right)$$

For the whole board, with $\mathcal{S}$ as the number of slot on the board, the formula would become :

$$H(\text{Whole board}) = \mathcal{S} * H(\mathcal{X}_i) = -\log_2\left(\frac{1}{\mathcal{C}^{\mathcal{S}}}\right)$$

The log in base 2 is larger in absolute value for smaller values of probability. Thus, the higher the color number and the higher the slot number, the higher the entropy is.

Question 16 : Optimized guessing

In order to do the best guess at each round, we need to maximize the entropy of the guess colour code. It will maximize the amount of information given by the colour.

One method to do that is to first put four pawns of the same colour (whatever the colour). By doing this, the information we collect is if the colour we first choose is present on the board one or several times.

Secondly, if this colour is correct or incorrect, we keep one or several pawn of this colour (we will slide it at each step to know its exact position). Then, we fill the empty slots with another colour (all the same). It allows to see if the second color is on the board. We continue this strategy by adding another colour till the end to know all the colours that are present. With this technique, the entropy is maximized at each step.

9

For instance, if the code to guess is 1 7 3 6 and our first guess is 2 2 2 2. The probability distribution looks like this :

Colours / Slot number	1	2	3	4
1	1/5	1/5	1/5	1/5
2	0	0	0	0
3	1/5	1/5	1/5	1/5
4	1/5	1/5	1/5	1/5
5	1/5	1/5	1/5	1/5
6	1/5	1/5	1/5	1/5
Entropy	2,32	2,32	2,32	2,32

The total entropy is then 9,28 which is pretty high. We know that 2 is not on the board. Then, if we do a second guess as 3 3 3 3, we obtain the followed result :

Colours / Slot number	1	2	3	4
1	1/4	1/4	0	1/4
2	0	0	0	0
3	0	1/4	1	1/4
4	1/4	1/4	0	1/4
5	1/4	1/4	0	1/4
6	1/4	1/4	0	1/4
Entropy	2	2	2	2

The total entropy at this step is then equal to 6 and still gives a lot of information for the next guess. If we continue this strategy, we will find the combination in a minimal of guesses.