Théorie de l'information

O. Kohel

18/01/2021

Rappel :

Théorème de Shannon (codage source)

① Soit $C : X \to A^*$ uniquement décodable.

Alors $\frac{H(X)}{\log_2 |A|} \leq E(\ell_C)$

② Il existe un codage u.d. $C : X \to A^*$ tel que

$$\frac{H(X)}{\log_2 |A|} \leq E(\ell_C) \leq \frac{H(X)}{\log_2 |A|} + 1.$$

Preuve (Idée). L'existence (②) suit du fait que le codage de Huffman satisfait $E(\ell_C) \leq H(X)/\log_2 |A| + 1$. La première inégalité suit du théorème de Kraft (qui suit). □

Théorème (Kraft–McMillan)

① Si $C : X \to A^*$ est u.d. avec fonction de longueur $\ell_C : X \to \mathbb{N}$, alors

$$\sum_{x \in X} q^{-\ell_C(x)} \leq 1, \quad (*)$$

où $q = |A|$.

(Kraft suite...)

② Si une fonction $l : X \longrightarrow \mathbb{N}^*$ satisfait ⊛ : $\sum_{x \in X} q^{-l(x)} \leq 1$, alors il existe un codage $C : X \longrightarrow A^*$ avec $l_C = l$.

Preuve. Exercices à venir.

Entropie conditionnelle.

Probabilité jointe et conditionnelle

Defn. Soit donné :

(i) Un espace de proba X $(= (X, p_X))$.

(ii) Un espace de proba Y $(= (Y, p_Y))$.

Une proba jointe sur $X \times Y$ est $p : X \times Y \longrightarrow [0,1]$ une proba. telle que

- $p_X(x) = p(\{x\} \times Y)$ et
- $p_Y(y) = p(X \times \{y\})$.

Defn. (suite) avec ces defns, l'information mutuelle $I(X,Y)$ de X et Y est

$$I(X,Y) = \sum_{(x,y)} p((x,y)) \log_2 \left(\frac{p((x,y))}{P_X(x) P_Y(y)} \right).$$

Remarque.

$I(X,Y)$ est un mesure du degré de dépendence (au sens probabiliste) de X et Y.

Si $p(x,y) = P_X(x) P_Y(y)$, alors $I(X,Y) = 0$.

On dit que la probabilité jointe $p((x,y)) = P_X(x) P_Y(y)$ est la probabilité produit.

Les données suivantes sont équivalentes (sur X et Y):

(i) Une probabilité jointe

$$p : X \times Y \longrightarrow [0,1].$$

(ii) Une probabilité conditionnelle $p(\ |x) : Y \longrightarrow [0,1]$, pour tout $x \in X$ tel que $p_X(x) \neq 0$, satisfaisant

$$\sum_{x \in X} p_X(x)\, p(y|x) = p_Y(y). \qquad (**)$$

(iii) Une probabilité conditionnelle $p(\ |y) : X \longrightarrow [0,1]$ pour tout $y \in Y$ tel que $p_Y(y) \neq 0$, satisfaisant

$$\sum_{y \in Y} p_Y(y)\, p(x|y) = p_X(x). \qquad (***)$$

Remarque. Dans (ii) (et (iii)) on peut définir p_Y (et p_X) par (**) (et (***)). Donc il suffit

(i) $p : X \times Y \longrightarrow [0,1]$ ($\Rightarrow p_X$ et p_Y).
(ii) p_X, $p(\ |x)$, et (iii) p_Y, $p(\ |y)$.

Preuve d'équivalence : Il suffit de détailler (i) $\Leftrightarrow$ (ii).

(i) $\Rightarrow$ (ii) : On définit :

a) $p_X : X \longrightarrow [0,1]$
$x \longmapsto p(\{x\} \times Y)$

et

$p_Y : Y \longrightarrow [0,1]$

b) $p(\ |x) : Y \longrightarrow [0,1]$ pour $p_X(x) \neq 0$
$y \longmapsto \dfrac{p((x,y))}{p_X(x)}$

On peut vérifier que p_X est une probabilité et $p(\ |x)$ aussi, satisfaisant

$$\sum_{x \in X} p_X(x)\, p(y|x) = \sum_{x \in X} p((x,y)) = p(X \times \{y\}) = p_Y(y).$$

(ii) $\Rightarrow$ (i) On pose

$$p((x,y)) = \begin{cases} p_X(x)\, p(y|x) \\ 0 \text{ si } p_X(x) = 0. \end{cases}$$

Défn L'entropie conditionnelle sur y, notée $H(X|y)$, est

$$H(X|y) = \sum_{x \in X} p(x|y) \log_2 (p(x|y)^{-1})$$

si $p(y) \neq 0$. L'entropie conditionnelle sur Y, notée $H(X|Y)$, est

$$H(X|Y) = \sum_{y \in Y} p_Y(y) H(X|y).$$

Avec ces définitions on peut définir :

Lemme. $I(X,Y) = H(X) - H(X|Y)$
$= H(Y) - H(Y|X)$.

N.B. $0 \overset{①}{\leqslant} H(Y|X) \overset{②}{\leqslant} H(Y)$

① Si on a égalité, on dit que Y est dépendent de X. ($I(X,Y) = H(Y)$)

② Si on a égalité, on ($\leqslant H(X)$) dit que Y est indép d'X. ($I(X,Y) = 0$)

Preuve (lemme) : Exercice.

Information mutuelle dans le contexte des codes correcteurs d'erreurs.

On pose $X = M =$ espace de messages

$Y = \mathcal{R} =$ espace de messages reçus

Idée :

$$M \xrightarrow[\text{bruyant}]{\text{canal}} \mathcal{R}$$

Le canal bruyant est modélisé par une conditionelle

$$p(\ |m) : \mathcal{R} \longrightarrow [0,1]$$

Pour un message transmis, on a une distribution de messages reçus.

Objectif : $I(X,Y) = H(X)$, ce qui signifie l'absence de perte d'information.

Information mutuelle dans le contexte de la cryptographie

On rappelle qu'un cryptosystème est une application

$$\mathcal{E} : \mathcal{K} \times \mathcal{M} \longrightarrow \mathcal{C},$$

où

$\mathcal{K}$ = espace de clés (souvent avec la proba uniforme)

$\mathcal{M}$ = espace de messages

$\mathcal{C}$ = espace de textes chiffrés

On dit aussi des textes clairs.

$X = \begin{cases} \mathcal{M} \text{ (avec } K \in \mathcal{K} \text{ fixée)} \\ \mathcal{K} \times \mathcal{M} \end{cases}$, $Y = \mathcal{C}$

Objectif. $I(\mathcal{K} \times \mathcal{M}, \mathcal{C}) = 0$

Pour un bon cryptosystème, la proba induite sur $\mathcal{C}$ doit être uniforme : $H(\mathcal{C}) = \log_2(|\mathcal{C}|)$.

Codes correcteurs d'erreurs.

On rappelle qu'un canal définit une proba conditionnelle sur $Y = \mathcal{R}$ pour tout $m \in X = \mathcal{M}$. On définit la capacité du canal par

$$C = \max_{p : \mathcal{M} \to [0,1]} (I(\mathcal{M}, \mathcal{R})).$$

Defn. Un canal est symétrique si $p(y|x)$ est indépendent du choix de couple (x,y) avec $y = x$.

Ici on suppose que $\mathcal{M} = \mathcal{R}$ en tant qu'ensembles.

Théorème (Shannon, Canal bruyant)
Soit donnés un canal bruyant avec capacité C et un $\varepsilon > 0$. Il existe un codage $C : \mathcal{M} \to \mathcal{R} \subseteq A^*$ avec rendement $R < C$ qui permet la transmission de l'info avec proba $< \varepsilon$ d'erreur.

Défn. Soit $C: X \to A^n$ un codage en bloc ($C(X) \subseteq A^n$), alors tous les mots de code sont de la même longueur), sans pertes. Le rendement R (ou taux de transmission) est $\frac{\log|X|}{n \log|A|}$.

Si $C: X \to A^*$ est un codage u.d. on définit

$$R = \limsup_{n \to \infty} \frac{\log|C^*(X^*) \cap A^n|}{n \log|A|}.$$

Cette deuxième définition s'accorde avec la première quand C est un codage en bloc (dans A^n).

Exemple $X = \{a, b, c\} \xrightarrow{C} \{0,1\}^*$

$C(a) = 0$, $C(b) = 10$, $C(c) = 11$.

$n=1$: $C^*(X^*) \cap A = \{0\}$ $\quad A = \{0,1\}$

$$\frac{\log_2(|C^*(X^*) \cap A|)}{\log_2|A|} = \frac{0}{1} = 0$$

$n=2$: $\{00, 10, 11\}$

$$\frac{\log_2(C^*(X^*) \cap A^2)}{2\log_2|A|} = \frac{\log_2 3}{2} > 1/2 \quad = 0{,}549\ldots$$

$n=3$: $|\{000, 010, 100, 011, 110\}| = 5$

aaa ab ba ac ca

$$\log_2(5)/3 \log_2 2 = \log_2 5/3 > 2/3$$

$n=4$: $|\{0000, 0010, 0100, 0100, 1000,$
$0011, 0110, 0110, 1100,$
$1010, 1011, 1110, 1111\}| = 13$

$$\log_2(13)/4 \log_2 2 = \log_2(13)/4 > 3/4 \quad = 0{,}925\ldots$$

En effet, on peut montrer que $\dfrac{\log_2(|C^*(X^*) \cap A^n|)}{n} \geqslant \dfrac{n-1}{n}$.

Donc le rendement est $R = 1$.

Le rendement mesure la proportion de caractères parmi n dans A^n qui porte de l'information (asymptotiquement).

Remarques. Le théorème de Shannon donne un borne sur le rendement pour lequel on peut espérer trouver un codage donnant une transmission fiable. Si le rendement dépasse la capacité, on peut pas éviter des pertes d'information. En pratique on cherche des bons codes avec rendement donné permettant

de réaliser une transmission fiable. Des codages en bloc, et les codages linéaires en particulier sont particulièrement intéressant en pratique. On va regarder des construction explicites dans la suite.