

Informatique II - Série 10

Exercice 10-1: La formule de Bayes

Pour un nombre d'exemples N suffisamment large, on peut estimer la probabilité conditionnelle $p(A|x)$ en prenant la formule

$$p(A|x) = \frac{n(x, A)}{N(x)} \quad (1)$$

où $N(x)$ est le nombre d'exemples avec une valeur x et $n(x, A)$ le nombre d'exemples avec des valeur x et A . Equation 1 devient exact dans la limit $N \rightarrow \infty$.

a) En prenant la formule 1 et des formules analogues pour $p(x)$ et $p(x, A)$ montrer que

$$p(x, A) = p(A|x) p(x) \quad (2)$$

b) Si il y a seulement deux 'classes' A et B, montrer que

$$p(x) = p(x, A) + p(x, B) \quad (3)$$

Exercice 10-2: Applications de la formule de Bayes

Il y a deux espèces d'ours, appelé 'P' et 'G'. Dans une population de 100 ours adults de l'espece 'P', 70 ont une taille de moins de 80 cm (petit) et 30 une taille entre 80 et 100 cm (moyen). Dans une population de 100 ours adults de l'espece 'G', 40 ont une taille entre 80 et 100 cm (moyen) et 60 une taille plus grande que 100 cm (grand).

a) Dessiner les distributions de taille pour les deux espèces.

b) Calculer $p(x, G)$ et $p(x, P)$ où x note la taille (x_p pour petit, x_m pour moyen, x_g pour grand). La taille de chaque population est de 100 ours exactement.

c) Vous observez un ours de taille x_g , c.a.d., plus grand que 100 cm. Quel est la probabilité qu'il s'agit de l'espece G ? Avant d'appliquer la formule, raisonner intuitivement.

d) Dans une région de montagne où chacune des deux espèces a une population de 20 ours, vous observez un ours de taille x_m , c.a.d., entre 80 et 100 cm. Quel est la probabilité qu'il s'agit de l'espece G ?

e) Dans une forêt où la population des ours P est 10 fois plus importante que celle des G , vous observez un ours de taille x_m , c.a.d., entre 80 et 100 cm. Quel est la probabilité qu'il s'agit de l'espece G ?

Exercice 10-3: Information Mutuelle: 'erasure channel'

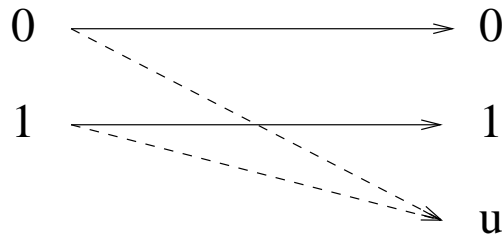
Des symbol $X = 0$ et $X = 1$ sont transmis à travers d'un canal bruité. Avec un probabilité $1 - p$ les symbols arrivent correctement, c.a.d., $Y = X \in \{0, 1\}$. Par contre, avec probabilité p l'information disparaît et est remplacée par u (unknown).

Exemple: une séquence

101011000100

peut se transformer lors de la transmission en

1010u10u0100



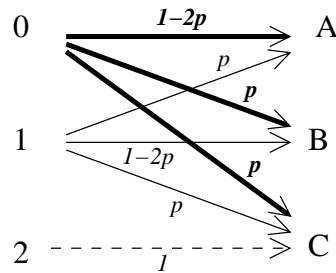
1. On suppose que les 2 symbols $x=0$ et $x=1$ sont utilisés avec la même probabilité. Calculer l'entropie H_0 de cette distribution des valeurs pour X .
2. Quelle est la probabilité $P(Y = 0)$ de recevoir un 0 ? Quelles sont les probabilités respectives $P(X = 0 | Y = 0)$ et $P(X = 1 | Y = 0)$? Quelle est enfin l'entropie du symbole émis $H(X | Y = 0)$?
3. De même si l'on reçoit un u ?
4. Calculer l'information mutuelle.

$$\begin{aligned}
 I &= H_0 - \langle H(X | Y) \rangle \\
 &= H_0 - P(Y = 0) H(X | Y = 0) - P(Y = 1) H(X | Y = 1) \\
 &\quad - P(Y = u) H(X | Y = u)
 \end{aligned}$$

Donner une valeur numérique pour le cas $p = 0.1$

Exercice 10-4: Information Mutuelle (finir à la maison)

On considère le canal bruité suivant:



Analytiquement, il se décrit de la manière suivante.

$$\begin{array}{lll}
 P(Y = A | X = 0) = 1 - 2p & P(Y = A | X = 1) = p & P(Y = A | X = 2) = 0 \\
 P(Y = B | X = 0) = p & P(Y = B | X = 1) = 1 - 2p & P(Y = B | X = 2) = 0 \\
 P(Y = C | X = 0) = p & P(Y = C | X = 1) = p & P(Y = C | X = 2) = 1
 \end{array}$$

On suppose que les trois symboles 0, 1 et 2 sont utilisés avec la même probabilité $P(X = 0) = P(X = 1) = P(X = 2) = \frac{1}{3}$.

1. Calculer l'entropie H_0 de cette distribution des valeurs pour X .
2. Quelle est la probabilité $P(Y = A)$ de recevoir un A ? Quelles sont les probabilités respectives $P(X = 0 | Y = A)$, $P(X = 1 | Y = A)$ et $P(X = 2 | Y = A)$? Quelle est enfin l'entropie du symbole émis $H(X | Y = A)$?
3. De même si l'on reçoit un C ?
4. Calculer l'information mutuelle.

$$\begin{aligned}
 I &= H_0 - \langle H(X | Y) \rangle \\
 &= H_0 - P(Y = A) H(X | Y = A) - P(Y = B) H(X | Y = B) \\
 &\quad - P(Y = C) H(X | Y = C)
 \end{aligned}$$

Donner une valeur numérique pour le cas $p = 0.1$

5. Peut-on augmenter l'information mutuelle en utilisant plus souvent le symbole C ? Expliquer pourquoi.

Exercice 10-5 Information mutuelle (à la maison)

On considère un canal bruité binaire. On note X le bit en entrée, X' le bit reçu après transmission. On fait l'hypothèse que le bit en entrée a une probabilité $1/2$ d'être égal à 1, donc une probabilité $1/2$ d'être égal à 0. La transmission s'effectue avec un "flip noise" de probabilité ϵ .

1. Calculer $I(X, X')$
2. Calculer...
 - $P(X = 0)$
 - $P(X = 1)$
 - $P(X' = 0)$
 - $P(X' = 1)$
3. Calculer...
 - $P(X = 0, X' = 0)$
 - $P(X = 0, X' = 1)$
 - $P(X = 1, X' = 0)$
 - $P(X = 1, X' = 1)$
4. Calculer $H(X, X') = - \sum_{X \in \{0,1\}} \sum_{X' \in \{0,1\}} p(X, X') \log p(X, X')$
5. Calculer $H(X) + H(X') - H(X, X')$
6. Conclure en comparant avec le résultat vu dans le cours.

Le but de cet exercice est de montrer que l'information I peut se calculer de deux manières différentes:

(a) $I = H_0 - \langle H(X | X') \rangle$
 $= H_0 - P(X' = 0) H(X | x' = 0) - P(X' = 1) H(X | x' = 1)$ vu dans le cours.

(b) $I = H(X) + H(X') - H(X, X')$

Cette équivalence sera également utilisée dans un des mini-projets.