

Section	1 ING IDSD
Matière	TP Théorie de l'information et de l'incertain
Enseignante	Trabelsi Nessrine

TP 2: Entropie conditionnelle et Information mutuelle

Objectifs du TP:

- Se familiariser avec les notions “entropie conditionnelle” et “information mutuelle”
- Calcul de l'entropie conditionnelle élémentaire et de l'entropie conditionnelle moyenne
- Calcul de l'information mutuelle
- Choix des critères utiles pour la classification

Exercice

Le but de cet exercice est d'écrire un programme nous aidant à la prise de décision en évaluant les différents critères.

Pour tester le code, nous prenons cet exemple :

Ciel	Temperature	Humidite	Vent	JouerAuTennis
Ensoleille	Chaude	Elevee	Faible	Non
Ensoleille	Chaude	Elevee	Fort	Non
Couvert	Chaude	Elevee	Faible	Oui
Pluie	Tiede	Elevee	Faible	Oui
Pluie	Fraiche	Normale	Faible	Oui
Pluie	Fraiche	Normale	Fort	Non
Couvert	Fraiche	Normale	Fort	Oui
Ensoleille	Tiede	Elevee	Faible	Non
Ensoleille	Fraiche	Normale	Faible	Oui
Pluie	Tiede	Normale	Faible	Oui
Ensoleille	Tiede	Normale	Fort	Oui
Couvert	Tiede	Elevee	Fort	Oui
Couvert	Chaude	Normale	Faible	Oui
Pluie	Tiede	Elevee	Fort	Non

1. Ouvrir le fichier “entropie_conditionnelle_information_mutuelle - test.py” se trouvant sur le bureau.
2. Compléter le code avec le chemin du fichier “tennisdata.csv” pour lire et formater les données d’entrée.
3. Compléter la définition de ces fonctions:
 - a. *probability_xk_sachant_ym* (*table_source_X*, *table_source_Y*, *xk*, *ym*)
 en appelant la fonction *probability_xk*(*table_source*, *xk*)
 où *table_source_X* et *table_source_Y* sont deux tables à une seule dimension contenant chacune les valeurs d’une source (X ou Y) (ce qui correspond par exemple à une colonne du tableau d’en haut sans la première ligne contenant le nom de la source)
xk est un événement de la source X, *ym* est un événement de la source Y
 - b. *probability_sourceX_sachant_ym* (*table_source_X*, *table_source_Y*, *ym*)
 en appelant la fonction *probability_xk_sachant_ym* (*table_source_X*, *table_source_Y*, *xk*, *ym*)
4. En appelant les bonnes fonctions, calculer et afficher le résultat de:
 - a. $P(\text{Temperature}=\text{Tiede})$
 - b. Les probabilités de tous les événements de la source “Vent”
 - c. Probabilités de (JouerAuTennis / Ciel = Ensoleille)
5. Compléter la définition de ces fonctions:
 - a. *entropie_sourceX_sachant_ym* (*table_source_X*, *table_source_Y*, *ym*)
 On utilisera la fonction “entropy” de scipy.
 - b. *entropie_sourceX_sachant_sourceY* (*table_source_X*, *table_source_Y*)
 en appelant les fonctions *entropie_sourceX_sachant_ym* (*table_source_X*, *table_source_Y*, *ym*) et *probability_sourceX* (*table_source*)
6. En appelant les bonnes fonctions, calculer et afficher le résultat de:
 - a. L’entropie de la source principale $H(\text{JouerAuTennis})$
 - b. Les entropies conditionnelles élémentaires
 $H(\text{JouerAuTennis}/\text{Ciel}=\{\text{Ensoleille}, \text{Pluie}, \text{Couvert}\})$,
 $H(\text{JouerAuTennis}/\text{Temperature}=\{\text{Chaude}, \text{Tiede}, \text{Fraiche}\})$
 $H(\text{JouerAuTennis}/\text{Humidite}=\{\text{Elevee}, \text{Normale}\})$
 $H(\text{JouerAuTennis}/\text{Vent}=\{\text{Faible}, \text{Fort}\})$
 - c. Les entropies conditionnelles moyennes $H(\text{JouerAuTennis}/\text{Ciel})$,
 $H(\text{JouerAuTennis}/\text{Temperature})$, $H(\text{JouerAuTennis}/\text{Humidite})$ et
 $H(\text{JouerAuTennis}/\text{Vent})$
7. Quel critère permet le mieux de diminuer l’entropie de la source principale?
8. Compléter la définition de cette fonction:
information_mutuelle_sourceX_and_sourceY (*table_source_X*, *table_source_Y*)
 en appelant les fonctions *probability_xk_and_ym* (*table_source_X*, *table_source_Y*, *xk*, *ym*) et *information_mutuelle_xk_and_ym* (*table_source_X*, *table_source_Y*, *xk*, *ym*)
9. Calculer $I(\text{JouerAuTennis}; \text{Ciel})$, $H(\text{JouerAuTennis}; \text{Temperature})$,
 $H(\text{JouerAuTennis}; \text{Humidite})$ et $H(\text{JouerAuTennis}; \text{Vent})$