

UNIVERSITÉ DE SHERBROOKE
Département d'informatique

IFT 615
Intelligence artificielle

Quiz 1
Hiver 2024

Publié le jeudi 1^{er} février

À remettre le jeudi 15 février à minuit par turn-in Quiz1

PROFESSEUR

Froduald Kabanza

AUXILIAIRES D'ENSEIGNEMENT

D'Jeff Kanda Nkashama

djeff.nkashama.kanda@usherbrooke.ca

Léo Chartrand

leo.chartrand@usherbrooke.ca

INSTRUCTIONS

Ce quiz a quatre pages numérotées, y compris cette page de garde, avec des questions comptant pour 3 points de la note totale du cours.

Écrivez votre nom, prénom et matricule ci-dessous. Vous devez inscrire aussi votre nom et prénom en haut de chacune des autres pages à l'endroit indiqué.

NOM : _____

PRÉNOM : _____

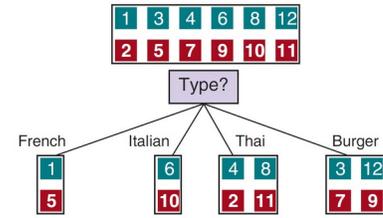
MATRICULE : _____

Questions – Arbre de décision et optimisation de modèles

1. **(0.5 point)** Encerchez les affirmations correctes. Les réponses à cette question seront évaluées en faisant une somme des points obtenus sur les affirmations encadrées correctement et en soustrayant les points obtenus sur les affirmations encadrées fautivement. Une différence négative sera ramenée à 0. Une différence dépassant 1 sera ramenée à 1. Il n'est pas nécessaire de savoir combien de points une affirmation correcte vaut ni le nombre d'affirmations correctes qu'il y a en tout.
- a. Les arbres de décision sont faciles à interpréter, car les règles de prédiction sont représentées par une série d'instructions de branchement simples.
 - b. La profondeur maximale d'un arbre de décision appris par l'algorithme LEARN-DECISION-TREE vu en classe, à partir de n données à k attributs catégoriques (donc k est la taille du vecteur d'entrée) est de $n*k$.
 - c. Les arbres de décision ne peuvent pas être utilisés pour générer un ensemble de règles "si-alors" qui peuvent être facilement comprises par les experts du domaine.
 - d. La profondeur maximale d'un arbre de décision appris à partir de n données à k variables (attributs) catégoriques (donc k est la taille du vecteur d'entrée) est de $k*m$ où m est l'entropie maximale pour une variable donnée.
 - e. La profondeur maximale d'un arbre de décision appris par l'algorithme LEARN-DECISION-TREE vu en classe, à partir de n données à k attributs catégoriques (donc k est la taille du vecteur d'entrée) n'est pas déterminée parce qu'elle dépend du choix de l'attribut à chaque étape.
 - f. Un arbre de décision peut être utilisé pour la classification et la régression.
 - g. Un arbre de décision peut être utilisé pour seulement pour la classification, mais pas pour la régression.
 - h. Un arbre de décision peut classifier des données non linéairement séparables.
 - i. L'algorithme *random forest* est une forme d'apprentissage ensembliste (*ensemble learning*) avec plusieurs arbres de décision en utilisant une technique de *bagging*.
 - j. L'algorithme *random forest* n'est pas utilisé en pratique dans l'industrie parce que les arbres de décision sont rendus désuets au profit des réseaux de neurones qui les ont remplacés dans
 - k. Un arbre de décision de profondeur 1 (un seul nœud décision) peut apprendre la fonction XOR (OU exclusif).
 - l. L'algorithme d'apprentissage des arbres de décision vu en classe (LEARN-DECISION-TREE) va dans le pire des cas produire un arbre contenant un nombre de nœuds intérieurs supérieur nombre d'attributs dans les exemples d'entraînement.
 - m. L'algorithme d'apprentissage des arbres de décision vu en classe (LEARN-DECISION-TREE) va dans le pire des cas produire un arbre de décision contenant un nombre de nœuds intérieurs égal au nombre d'attributs dans les exemples d'entraînement.
 - n. L'algorithme *random forest* n'est pas utilisé en pratique dans l'industrie parce que les arbres de décision sont rendus désuets au profit des réseaux de neurones qui les ont remplacés dans quasiment toutes les applications que l'on puisse imaginer.

2. (0.5 point) Soit le jeu de données et l'arbre de décision ci-après.

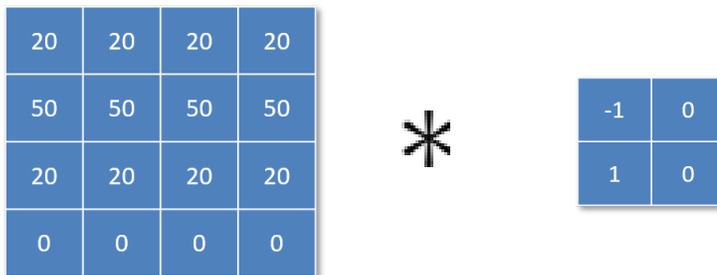
Example	Input Attributes										Output
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
x ₁	Yes	No	No	Yes	Some	\$\$\$	No	Yes	French	0-10	y ₁ = Yes
x ₂	Yes	No	No	Yes	Full	\$	No	No	Thai	30-60	y ₂ = No
x ₃	No	Yes	No	No	Some	\$	No	No	Burger	0-10	y ₃ = Yes
x ₄	Yes	No	Yes	Yes	Full	\$	Yes	No	Thai	10-30	y ₄ = Yes
x ₅	Yes	No	Yes	No	Full	\$\$\$	No	Yes	French	>60	y ₅ = No
x ₆	No	Yes	No	Yes	Some	\$\$	Yes	Yes	Italian	0-10	y ₆ = Yes
x ₇	No	Yes	No	No	None	\$	Yes	No	Burger	0-10	y ₇ = No
x ₈	No	No	No	Yes	Some	\$\$	Yes	Yes	Thai	0-10	y ₈ = Yes
x ₉	No	Yes	Yes	No	Full	\$	Yes	No	Burger	>60	y ₉ = No
x ₁₀	Yes	Yes	Yes	Yes	Full	\$\$\$	No	Yes	Italian	10-30	y ₁₀ = No
x ₁₁	No	No	No	No	None	\$	No	No	Thai	0-10	y ₁₁ = No
x ₁₂	Yes	Yes	Yes	Yes	Full	\$	No	No	Burger	30-60	y ₁₂ = Yes



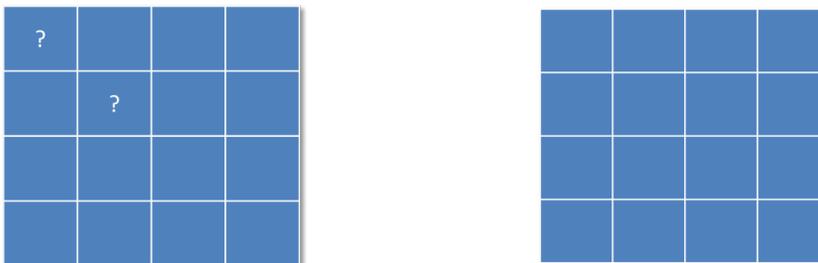
Calculez le gain d'information apporté par le choix de l'attribut *Pat* au nœud $\{3, 12, 7, 9\} \equiv \{x_3, x_{12}, x_7, x_9\}$. Mettez votre réponse dans l'encadrez ci-après.

Questions sur les réseaux de neurones à convolution

3. (0.5 points) Soit la matrice 4 x 4 et le filtre de convolution 2 x 2 ci-après.



Indiquez la valeur des entrées indiquées par « ? » ci-après dans la matrice résultant de l'application de la convolution avec un pas de palayage de 1. Indiquez vos valeurs dans la matrice vide à droite.



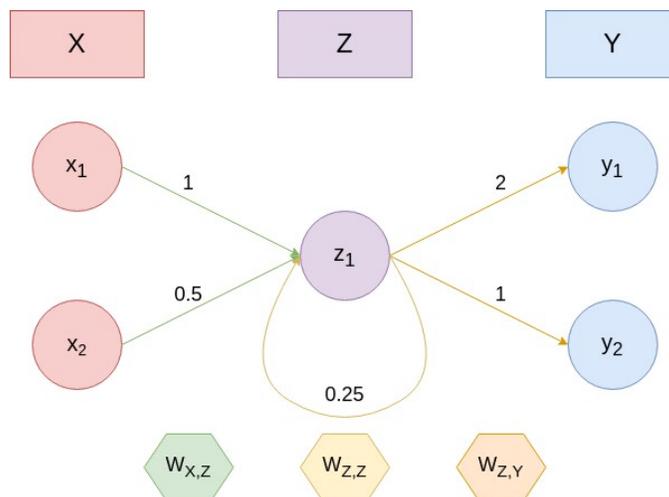
4. (0.5 points) Soit la matrice 4 x 4 ci-après.

20	20	20	20
50	50	50	50
20	20	20	20
0	0	0	0

Indiquez ci-après le résultat de l'application d'un max-pooling avec un filtre de taille 2 x 2 à cette matrice. Mettez votre réponse dans l'encadré ci-après

Questions sur les réseaux de neurones récurrents

5. (0.5 points) Soit le réseau de neurones récurrent ci-après:



Soit une séquence de texte $X = \{x_1, x_2, \dots, x_N\}$ projetée dans un espace d'embedding à 2 dimensions, où chaque élément x_i devient le vecteur

$$\langle x_{i,1}, x_{i,2} \rangle.$$

Après la transformation par la couche d'embedding, le premier élément x_1 devient le vecteur

$$\langle x_{1,1}, x_{1,2} \rangle = \langle 4, 6 \rangle.$$

Ce vecteur est ensuite interprété par le réseau de neurones récurrent illustré plus haut, où l'état caché z est initialement nul (i.e. $z_{0,1} = 0$). Le réseau ne comporte pas de biais ni de fonction d'activation. En tenant compte du fait qu'il s'agit du **premier élément de la séquence**, quelle est la sortie $y_1: \langle y_{1,1}, y_{1,2} \rangle$?

Mettez votre réponse dans l'encadrez ci-après:

6. **(0.5 points)** Après la transformation par la couche d'embedding, le deuxième élément x_2 devient le vecteur

$$\langle x_{2,1}, x_{2,2} \rangle = \langle -2, 5 \rangle.$$

En tenant compte du fait qu'il s'agit du **deuxième** élément de la séquence (pensez à l'état caché), quelle est la sortie $y_2: \langle y_{2,1}, y_{2,2} \rangle$?

Mettez votre réponse dans l'encadrez ci-après:

FIN DU QUIZ