UNIVERSITÉ DE SHERBROOKE
Département d'informatique

IFT 615
Intelligence artificielle

Examen périodique
Hiver 2024

PROFESSEUR
Froduald Kabanza

SOLUTIONS

Réservé. N'inscrivez rien dans ce tableau

	Q1 / 3	Q2 / 2	Q3 / 2	Q4 / 1	Q5 / 2	Q6 / 2	Q7 / 3	Total / 15
Note								

Question 1 (3 points) – Connaissances générales sur l'apprentissage automatique

Cochez les affirmations correctes, en supposant qu'il n'y a aucune contrainte indiquée sur l'architecture (par exemple, le nombre de neurones, la fonction d'activation) autre que la contrainte explicitement mentionnée dans la question. Un choix incorrect annule un choix correct. S'il y a autant ou plus de choix incorrects que corrects, le résultat sera zéro pour la sous-question.

a. (1 point) La fonction *xor(x,y)* représentée par la table de vérité de droite peut être apprise par :

x	y	$x \oplus y$
0	0	0
0	1	1
1	0	1
1	1	0

- □ i. Un réseau de neurones avec une seule couche (pas de couche cachée)
- ■ ii. Un réseau de neurones avec deux couches (une seule couche cachée)
- □ iii. Un arbre de décision de profondeur 1.
- ■ iv. Un arbre de décision de profondeur 2.

b. (1 point) La fonction *f(x,y)*= ┐*(x* ∨ *y)*, représentée par la table de vérité de droite peut être apprise par :

x	y	$\neg(x \vee y)$
0	0	1
0	1	0
1	0	0
1	1	0

- ■ i. Un réseau de neurones avec une seule couche (pas de couche cachée)
- ■ ii. Un réseau de neurones avec deux couches (une seule couche cachée)
- □ iii. Un arbre de décision de profondeur 1.
- □ iv. Aucune de ces trois dernières réponses

c. (1 point) Vous êtes en train d'entraîner un réseau de neurones pour une application quelconque (ex. : reconnaissance d'images). Vous constatez que la perte sur les données d'entraînement est devenue très basse, mais par contre la perte sur les données de validation demeure très élevée. Vous pourriez améliorer la perte sur les données de validation en :

- □ i. Augmentant le nombre de couches.
- ■ ii. Diminuant le nombre de couches.
- ■ iii. Diminuant la taille de la couche cachée.
- □ iv. Augmentant la taille de la couche cachée.

Question 2 (2 points) – Rétropropagation du gradient

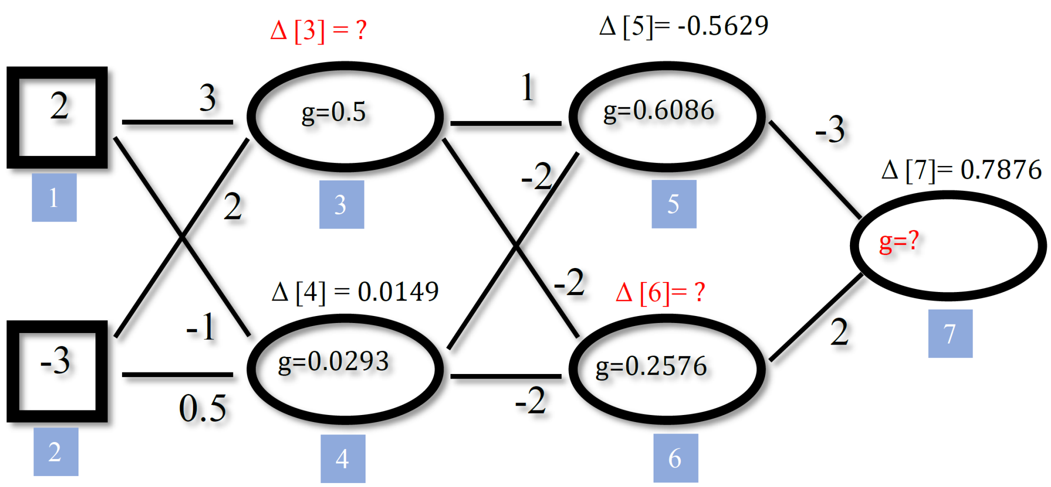
Soit le réseau de neurones ci-après, où $x = [2, -3]$ est la donnée fournie au réseau et $y = 1$ sa sortie désirée. Les valeurs dans les ellipses représentent la sortie du neurone correspondant lors application de la propagation avec la sigmoïde comme fonction d'activation.

Veuillez compléter les valeurs manquantes, indiquées par un « ? », pour la propagation-avant et les valeurs $\Delta[j]$ pour la rétropropagation.

Rappel des formules à utiliser :

$$\sigma(x\,;\,w) = \frac{1}{1 + e^{-w \cdot x}}$$

$$\Delta[j] = g(in_j)(1 - g(in_j)) \sum_k w_{j,k} \Delta[k]$$



$g[7] = 0.2124$

$\Delta[6] = 0.2576 * (1 - 0.2576) * [\,2 * 0.7876] = 0.3012$

$\Delta[3] = 0.5 * (1 - 0.5) * [-2 * 0.3012 + 1 * (-0.5629)] = -0.2913$

Question 3 (2 points) – Applications d'une convolution et d'un *max-pooling*

a. (**1 point**) Soit la matrice 4 x 4 et le filtre de convolution 2 x 2 ci-après.

20	0	10	30
0	10	30	0
0	40	20	10
50	10	0	20

0.5	-1
0	0.5

Indiquez la valeur des cellules blanches ci-après dans la matrice résultant de l'application de la convolution avec un pas de balayage de 1.

15		
	0	

b. (**1 point**) Soit la matrice 4 x 4 ci-après.

20	0	10	30
0	10	30	0
0	40	20	10
50	10	0	20

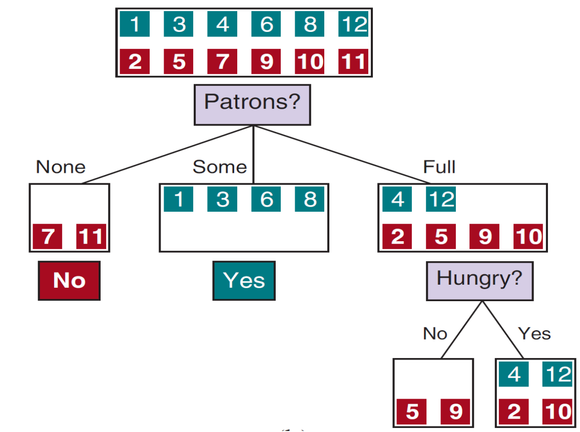
Indiquez ci-après le résultat de l'application d'un max-pooling avec un filtre de taille 2 x 2 à cette matrice. Mettez votre réponse dans l'encadré ci-après :

20	30
50	20

Question 4 (1 point) – Apprentissage des arbres de décision

(1 point) Soit le jeu de données et l'arbre de décision ci-après . Dans l'arbre de décision, les numéros correspondent aux exemples dans la table. Par exemple, 1 pour x1, 2 pour x2, ainsi de suite.

Example	Input Attributes										Output
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
$\mathbf{x}_1$	*Yes*	*No*	*No*	*Yes*	*Some*	*\$\$\$*	*No*	*Yes*	*French*	*0–10*	$y_1 = Yes$
$\mathbf{x}_2$	*Yes*	*No*	*No*	*Yes*	*Full*	*\$*	*No*	*No*	*Thai*	*30–60*	$y_2 = No$
$\mathbf{x}_3$	*No*	*Yes*	*No*	*No*	*Some*	*\$*	*No*	*No*	*Burger*	*0–10*	$y_3 = Yes$
$\mathbf{x}_4$	*Yes*	*No*	*Yes*	*Yes*	*Full*	*\$*	*Yes*	*No*	*Thai*	*10–30*	$y_4 = Yes$
$\mathbf{x}_5$	*Yes*	*No*	*Yes*	*No*	*Full*	*\$\$\$*	*No*	*Yes*	*French*	*>60*	$y_5 = No$
$\mathbf{x}_6$	*No*	*Yes*	*No*	*Yes*	*Some*	*\$\$*	*Yes*	*Yes*	*Italian*	*0–10*	$y_6 = Yes$
$\mathbf{x}_7$	*No*	*Yes*	*No*	*No*	*None*	*\$*	*Yes*	*No*	*Burger*	*0–10*	$y_7 = No$
$\mathbf{x}_8$	*No*	*No*	*No*	*Yes*	*Some*	*\$\$*	*Yes*	*Yes*	*Thai*	*0–10*	$y_8 = Yes$
$\mathbf{x}_9$	*No*	*Yes*	*Yes*	*No*	*Full*	*\$*	*Yes*	*No*	*Burger*	*>60*	$y_9 = No$
$\mathbf{x}_{10}$	*Yes*	*Yes*	*Yes*	*Yes*	*Full*	*\$\$\$*	*No*	*Yes*	*Italian*	*10–30*	$y_{10} = No$
$\mathbf{x}_{11}$	*No*	*No*	*No*	*No*	*None*	*\$*	*No*	*No*	*Thai*	*0–10*	$y_{11} = No$
$\mathbf{x}_{12}$	*Yes*	*Yes*	*Yes*	*Yes*	*Full*	*\$*	*No*	*No*	*Burger*	*30–60*	$y_{12} = Yes$



Calculez, à trois décimales près, la valeur du gain d'information apporté par le choix de l'attribut *Res* au nœud *{4,2,12,10}*.

- $B(q) = -(q \log_2 q + (1-q)\log_2(1-q))$
- $E(Sortie) = B(\frac{p}{p+n})$
- $Remainder(A) = \sum_{k=1}^{d} (\frac{p_k + n_k}{p+n})\ B(\frac{p_k}{p_k + n_k})$
- $Gain\ (A) = B(\frac{p}{p+n})$ - $Remainder(A)$

Yes No

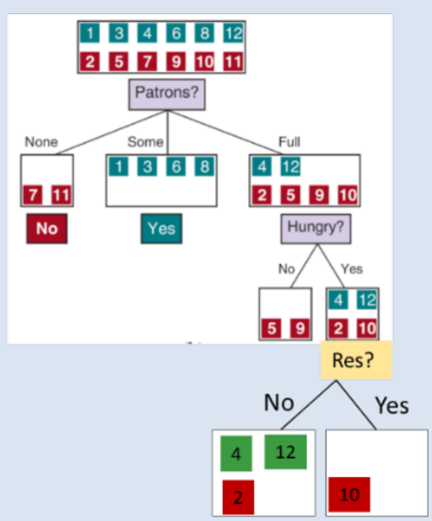
- $Gain\ (Res?)$

$= 1 - [\frac{1}{4} B(\frac{0}{1}) + \frac{3}{4} B(\frac{2}{3})]$

$= 1 - [0.25 * 0 + 0.75 * 0.91829]$

$= 1 - 0.688722$

≈ 0.31127



Question 5 (2 points) – Classifieur bayésien naïf et le modèle *n-gram*

Vrai ou Faux - Encerclez les affirmations correctes. Un choix incorrect annule un choix correct. S'il y a autant ou plus de choix incorrects que corrects, le résultat sera zéro pour la sous-question.

a. Le modèle *bag-of-words* et le modèle *unigram* ($n = 1$) sont essentiellement équivalents.

b. Le classifieur bayésien naïf est appelé « naïf » parce qu'on ne suppose pas l'indépendance conditionnelle.

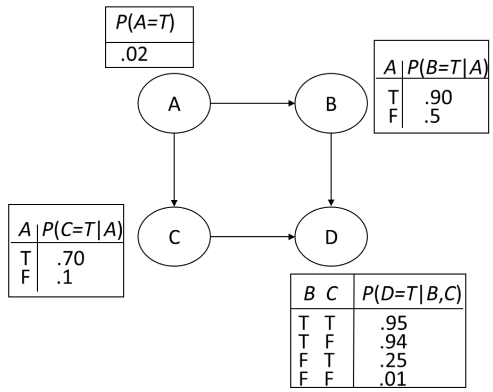
c. Lorsqu'on utilise un modèle *bag-of-words* pour faire de la classification, la fréquence d'un mot dans une séquence de texte influence la prédiction de classe.

d. Le modèle *n-gram* a une taille maximale de $n = 3$ (*trigram*).

e. Le modèle *bag-of-words* est particulièrement utile pour la génération de texte où la syntaxe est importante.

Question 6 (2 points) – Réseaux bayésiens

Soit le réseau bayésien (RB) suivant pour quatre (4) variables booléennes aléatoires (A, B, C et D) avec les tables de probabilités conditionnelles correspondantes.



f. **(1 point)** Donnez l'équation la plus simple possible pour calculer P(A=F | B= T, C=T, D=F) en fonction des distributions de probabilités conditionnelles du réseau bayésien.

$$P(A \mid B, C, D) = P(A,B,C,D) / P(B, C, D)$$

$$= P(D \mid B,C)\, P(B|A)\, P(C|A)\, P(A) \;/\; P(D \mid B, C)\, P(B,C)$$

$$= P(B|A)\, P(C|A)\, P(A) \;/\; \Sigma_a\, P(B \mid a)\, P(C \mid a)\, P(a)$$

g. **(1 point)** Donnez le résultat à deux décimales près du calcul

$$P(\neg a \mid b, c, \neg d) = P(b \mid \neg a)\, P(c \mid \neg a)\, P(\neg a) / [\; P(b \mid a)\, P(c \mid a)\, P(a) + P(b \mid \neg a)\, P(c \mid \neg a)\, P(\neg a) \;]$$

$$=0.5*0.1*0.98/[0.9*0.7*0.02+0.5*0.1*0.98]$$

$$=0.795$$

Question 6 (3 points) – Modèle de Markov caché

Soit un modèle de Markov caché d'ordre 1 dont les variables cachées H_t et les variables observées O_t ont toutes comme domaine les symboles *0 et 1*. Soit les distributions suivantes :

Modèle de transition		
	$H_{t-1} = 0$	$H_{t-1} = 1$
$P(H_t = 0 \mid H_{t-1})$	0.4	0.7
$P(H_t = 1 \mid H_{t-1})$	0.6	0.3

Modèle d'observation		
	$H_t = 0$	$H_t = 1$
$P(O_t = 0 \mid H_t)$	0.8	0.1
$P(O_t = 1 \mid H_t)$	0.2	0.9

Distribution initiale		
	$H_1 = 0$	$H_1 = 1$
$P(H_1)$	0.6	0.4

Supposez que l'on observe $O_1 = 0$, $O_2 = 1$ et $O_3 = 0$.

a. **(1 point)** Remplissez les **deux** entrées manquantes du tableau α ci-après, à quatre (4) décimales près.

$\alpha(i,t)$	$t = 1$	$t = 2$	$t = 3$
$i = 0$	0.48	0.044	0.16528
$i = 1$	0.04	0.27	0.01074

b. **(1 point)** Donnez la probabilité, à trois décimales près, d'observer *0*, suivi de *1* et suivi de *0*.

$P(S_1=0, S_2=1, S_3=0) = \text{å}_i\, \alpha(i,3) = 0.16528 + 0.01074 = \mathbf{0.17602}$

c. **(1 point)** Donnez la probabilité, à trois décimales près, que l'état caché au temps $t=3$ soit *1*, sachant que l'on a observé *0*, suivi de *1* et suivi de *0*.

$P(H_3 = 0 \mid S_1=0, S_2=1, S_3=0) = P(H_3 = 0, O_1=0, O_2=1, O_3=0) / \text{å}_i\, P(H_3 = i, O_1=0, O_2=1, O_3=0)$

$= \alpha(1,3) / \text{å}_i\, \alpha(i,3) = 0.01074 / 0.17602 = \mathbf{0.06102}$

FIN DE L'EXAMEN