

ATTENTION:

Quelques sujets n'ont pas été vus cette année:

1. Apprentissage supervisé: arbres de décision
2. Représentation des connaissances: Cadres (*Frames*)
3. Architectures d'apprentissage profond: Application détaillée de CNN

Quelques sujets ont été ajoutés cette année:

1. Logique floue
2. Apprentissage supervisé: SVMs
3. Apprentissage par renforcement

La réponse à certaines questions est laissée vide intentionnellement.

Certaines questions V/F seront remplacées par des questions de terminologie cette année.



CSI4506 - INTELLIGENCE ARTIFICIELLE

Examen final

Nom de l'étudiant: _____

Signature: _____

Numéro d'étudiant: _____

Professeure: Caroline Barrière

Date: le 20 décembre 2018

Durée: 180 minutes. Évaluation sur 100 points.

Examen à livre fermé. Calculatrice requise. AUCUN téléphone accepté.

Instructions: Répondez aux questions dans l'espace alloué dans le questionnaire. Soyez aussi explicite que possible. Si vous faites des hypothèses pour des situations que vous trouvez ambiguës, écrivez ces hypothèses.

BONNE CHANCE!

Question		Mark	Max	Question		Mark	Max
1	Recherches (V/F)		8	5b	TAL - Pipeline (algo)		12
2	Problèmes de Satisfaction de Contraintes (V/F)		8	6a	Représentation des connaissances (design)		10
3	Arbres de décision / Naive Bayes (V/F)		8	6b	RC - BOW overlap (algo)		5
4a	Réseaux de neurones (V/F)		10	7a	Architectures d'apprentissage profond (V/F)		6
4b	Réseaux de neurones (algo)		15	7b	Réseau convolutif (algo)		9
5a	Traitement Automatique des Langues (V/F)		8	8	Projet favori		1
			Total				100

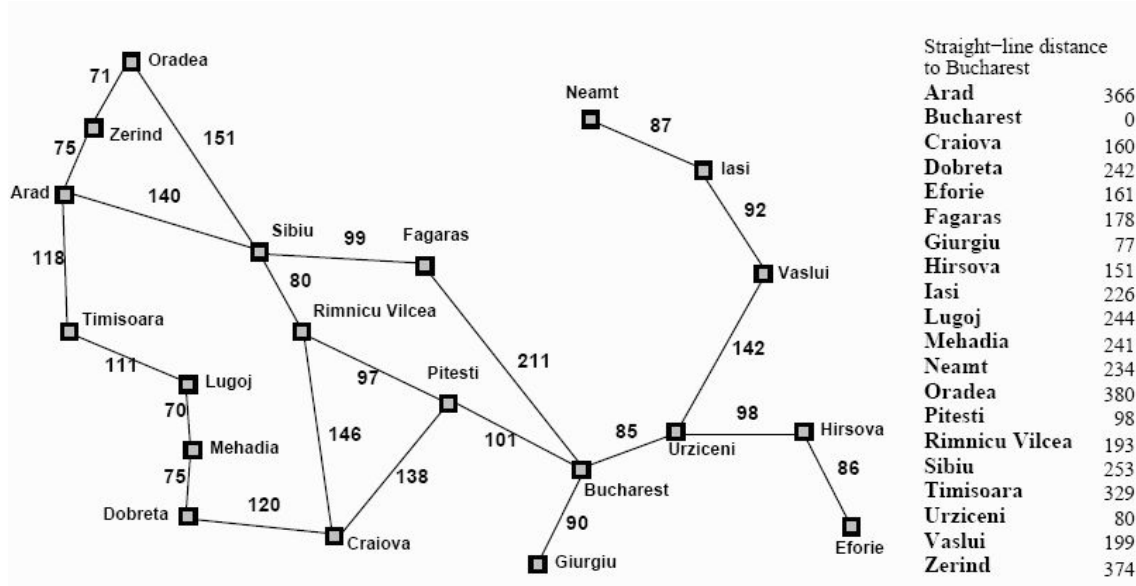
PARTIE 1 - Recherches aveugles et heuristiques (8 points)

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Dans un espace de recherche de solutions, le facteur de ramification (<i>branching factor</i>) correspond au nombre d'étapes effectuées par l'algorithme de recherche.	F
2	La recherche par approfondissement itératif (<i>iterative deepening search</i>) est un raffinement de la recherche en largeur (<i>breadth-first search</i>).	F
3	La recherche en profondeur trouvera toujours la solution optimale plus rapidement que la recherche en largeur.	F
4	L'algorithme "plus-court-chemin-d'abord" (<i>lowest-cost-first</i>) trouvera toujours la solution optimale.	V

PARTIE 1 - Recherches aveugles et heuristiques (8 points) - SUITE

Les questions 5 à 8 réfèrent à la carte ci-dessous.



ATTENTION: La première lettre de chaque ville est utilisée pour son identification.

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
5	Pour aller de A à B, la recherche gloutonne (<i>greedy</i>) propose le chemin: A, Z, O, S, F, B.	F
6	Pour aller de S à B, la recherche A* propose le chemin: S, F, B.	F
7	Pour aller de S à B, la recherche "best-first search" propose le chemin: S, R, P, B.	F
8	Peu importe l'algorithme (<i>greedy</i> , <i>best-first</i> , A*), le chemin pour aller de N à B sera le même.	T

PARTIE 2 - Problèmes de satisfaction de contraintes (8 points)

Le tableau ci-bas montre les disponibilités de chaque professeur pour l'enseignement d'un cours. Le tableau indique l'heure minimale et l'heure maximale pour le début du cours. Nous supposons que tous les cours sont d'une durée d'une heure.

	Début Min	Début Max
Alban	15:00	17:00
Bo	13:00	16:00
Giu	14:00	15:00
Youla	13:00	16:00

Le but est d'assigner une heure de cours à chaque professeur en respectant les contraintes suivantes:

Bo doit enseigner avant Alban.

Giu doit enseigner avant Youla.

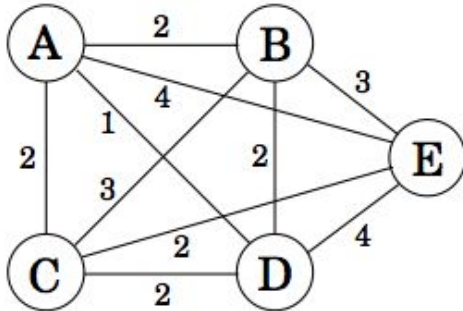
Youla doit enseigner avant Alban.

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Un algorithme "generate-and-test" devra tester 72 solutions.	V
2	Les domaines pour Alban et Bo sont cohérents.	V
3	Les domaines pour Giu et Youla sont cohérents.	F
4	Le problème ci-haut ne contient pas de contraintes souples.	V

PARTIE 2 - Problèmes de satisfaction de contraintes (8 points) - SUITE

Étant donné le graphe ci-bas pour les questions 5 à 8, qui réfère au le problème du voyageur de commerce (*travelling salesman problem*). Le but est de visiter toutes les villes de la façon la moins coûteuse.



Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
5	Supposons que notre solution peut débuter de n'importe quel noeud. Si nous utilisons l'algorithme glouton (<i>greedy</i>), n'importe quel noeud de départ générera une solution qui aura la même longueur totale pour le chemin parcouru.	V
6	Supposons une première solution (<i>greedy</i>): C, A, D, B, E. Il serait possible pour un algorithme de modification aléatoire (<i>random modification</i>) de suggérer le chemin C, D, A, B, E comme prochaine solution à évaluer.	V
7	Supposons une première solution (<i>greedy</i>): C, A, D, B, E. Si, par la suite, nous utilisons le recuit simulé (<i>simulated annealing</i>), et que celui-ci génère une solution C, A, B, D, E, alors cette solution sera toujours acceptée et l'algorithme poursuivra avec cette nouvelle solution.	F
8	Si nous choisissons un algorithme par étape aléatoire (<i>random step</i>), il ne sera pas possible d'explorer les noeuds dans l'ordre alphabétique A, B, C, D, E.	F

PARTIE 3 - Apprentissage machine supervisé (8 points)

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Si une pièce de monnaie est biaisée pour toujours tomber sur <i>face</i> , son contenu d'information (information content) est alors de 1.0.	F
2	Dans un arbre de décision, la racine de l'arbre a le plus faible contenu d'information.	F
3	Les arbres de décision sont des classificateurs binaires.	F
4	Dans les arbres de décision, le gain d'information est calculé pour classifier un exemple test (exemple ne faisant pas partie de l'ensemble d'entraînement).	F

PARTIE 3 - Apprentissage machine supervisé (8 points) - SUITE

Supposons l'ensemble d'entraînement suivant, qui servira pour les énoncés de 5 à 8. Il y a 3 attributs à l'entrée (durée, genre, rating) et une classe à prédire (appréciation du public).

	Attributs considérés			Classe à prédire
Film	Durée	Genre	Rating	Appréciation du public
M1	90	Drame	PG-13	Mauvais
M2	140	Drame	R	Bon
M3	80	Comedie	R	Mauvais
M4	90	Drame	G	Mauvais
M5	100	Drame	PG-13	Mauvais
M6	110	Comedie	R	Bon
M7	100	Comedie	G	Mauvais

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
5	La probabilité a priori de Mauvais est 5/7.	V
6	La probabilité a posteriori $P(\text{Mauvais} \text{R})$ est 2/7.	F
7	$P(\text{Mauvais} \text{Drame}) > P(\text{Bon} \text{Comedie})$	V
8	Étant donné cet ensemble d'entraînement, un nouveau film à classifier qui a comme attribut PG-13 sera toujours classifié comme Mauvais.	V

PARTIE 4a - Réseaux de neurones (10 points)

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Un perceptron contient une seule couche cachée.	F
2	L'entropie croisée est utilisée comme fonction d'erreur pour la régression linéaire.	F
3	L'apprentissage dans un réseau de neurones consiste à modifier les poids de façon itérative.	V
4	Dans un perceptron multi-couches, la couche cachée devrait contenir le même nombre de noeuds que la couche d'entrée.	F
5	Pour la régression logistique, la valeur de sortie est binaire.	V
6	Le terme "profond" dans "apprentissage profond" vient du fait qu'on empile plusieurs couches cachées dans un réseau de neurone.	V
7	Pour classifier des phrases en positif/négatif, un Perceptron serait approprié.	V
8	La sortie de la fonction sigmoïde varie entre 0 et 1.	V
9	La descente de gradient est un algorithme utilisé pour faire l'ajustement des poids dans un réseau de neurones.	V
10	Les poids sont initialisés à +1 et -1 au début de la descente de gradient pour l'apprentissage d'un réseau de neurones.	F
11	Une condition d'arrêt possible de la descente de gradient est d'obtenir une erreur trop grande.	F
12	Étant donné un problème de classification de crédit qui aurait 2 attributs, soit le salaire (moins de 50K, 50K à 100K, plus de 100K) et l'antécédent de crédit (bon, moyen, mauvais), nous aurions alors besoin de 5 noeuds d'entrée dans la couche d'entrée pour représenter cette information.	F
13	Un encodage des entrées de type "one-hot-encoding" réduira le nombre d'entrées.	F
14	Un modèle qui "overfit" sur l'ensemble d'entraînement aura probablement une moindre capacité de prédiction sur l'ensemble test.	V
15	Plus il y a de couches cachées dans un réseau de neurones, moins il est probable qu'il y ait du "overfitting" sur l'ensemble d'entraînement.	F

PARTIE 4b - Réseaux de neurones (15 points)

Étant donné le perceptron multi-couches ci-bas comprenant 2 unités d'entrée, 2 unités cachées et une unité de sortie.

Notation:

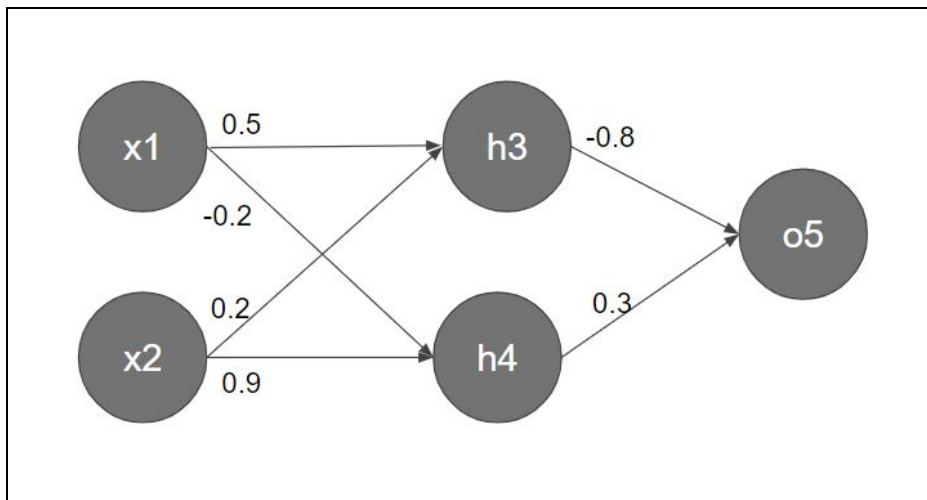
w_{ij} réfère au poids du noeud i au noeud j

net_i réfère au résultat de la partie linéaire de l'activation sur le noeud i

out_i réfère au résultat de la partie non-linéaire de l'activation sur le noeud i , soit $out_i = \text{sigmoid}(net_i)$.

Tel que montré dans la figure ci-bas, le réseau a les poids initiaux suivant:

$w_{13}=0.5$, $w_{14}= -0.2$, $w_{23} = 0.2$, $w_{24} = 0.9$, $w_{35} = -0.8$, $w_{45} = 0.3$.



1. Calculer la prédiction à la sortie out_5 , pour le réseau ci-haut, si l'entrée est $x1 = 1.0$, $x2 = 1.0$.
2. Si nous utilisons la somme des erreurs au carré comme fonction d'erreurs, calculer l'erreur à la sortie si la sortie désirée est de 1.0.
3. Faites le calcul du nouveau poids w_{14} en utilisant la rétro-propagation des erreurs. Supposez un taux d'apprentissage de 0.1. Assurez-vous d'écrire l'équation de rétro-propagation des erreurs clairement avant de faire les calculs. Les dérivées nécessaires sont fournies à la page suivante.

PARTIE 4b - Réseaux de neurones (15 points) - RÉPONSE

1)

$$\text{Net}_3 = 1 \cdot 0.5 + 1 \cdot 0.2 = 0.7$$

$$\text{Net}_r = 1 \cdot -0.2 + 1 \cdot 0.9 = 0.7$$

$$\text{Out}_3 = 1 / (1 + e^{-\text{net}_3}) = 0.668$$

$$\text{Out}_4 = 0.668$$

$$\text{Net}_5 = 0.668 \cdot -0.8 + 0.668 \cdot 0.3 = -0.334$$

$$\text{Out}_5 = \text{sigmoide}(\text{Net}_5) = 0.417$$

2) $(1.0 - 0.417)$ au carré = 0.34

3)

$$(\text{out}_5 - \text{target}_5) \cdot (\text{out}_5 \cdot (1 - \text{out}_5)) \cdot w_{45} \cdot (\text{out}_4 \cdot (1 - \text{out}_4)) \cdot x_1$$

$$(0.417 - 1) \cdot (0.417 \cdot (1 - 0.417)) \cdot 0.3 \cdot (0.668 \cdot (1 - 0.668)) \cdot 1 = -0.0094$$

$$W_{14}(t+1) = w_{14}(t) - 0.1 \cdot -0.0094 = -0.2 + 0.00094 = -0.19906$$

$$\frac{\partial SSE}{\partial out_j} = (out_j - target_j)$$

$$\frac{\partial out_j}{\partial net_j} = out_j \cdot (1 - out_j)$$

$$\frac{\partial net_j}{\partial out_i} = w_{ij}$$

$$\frac{\partial net_j}{\partial w_{ij}} = x_i$$

PARTIE 5a - Traitement automatique des langues (8 points)

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Le résumé automatique est une des applications possibles du traitement automatique des langues.	V
2	Il sera moins probable de retrouver des erreurs typographiques dans un article de journal que dans un courriel.	F
3	Un lemmatiseur permettra de réduire plusieurs formes de mots à une seule forme unique.	V
4	Un algorithme de "stemming" enlève des caractères au début des mots.	F
5	La variation entre le mot "beau" et le mot "beauté" est un exemple d'une variation d'inflection (<i>inflectional variation</i>).	F
6	Un étiqueteur de partie du discours (POS tagger) assignera à chaque entité une étiquette, soit (PER, ORG, LOC).	F
7	Le mots "cours" dans ces deux phrases: "Je <i>cours</i> tous les matins." et "J'ai un <i>cours</i> de mathématique le jeudi." est une illustration de l'ambiguïté dans l'assignation de parties du discours.	V
8	Pour identifier des numéros de téléphone dans un texte, une bonne approche serait d'utiliser des expressions régulières.	V
9	Pour la reconnaissance d'entités nommées, un gazetteer est utile pour les types d'entités dont l'ensemble est énumérable.	V
10	Les points (.) sont probablement de bons indicateurs de positions de séparation de phrases.	V

PARTIE 5b - Traitement automatique des langues (12 points)

Étant donné la phrase: “*I went to the movies at the renovated BlueBird Theater in Ottawa with my friends Ali and Yu.*” En supposant que la tokenization a déjà été faite (voir tableau) en ignorant les ponctuations, ajouter dans le tableau les résultats attendus des opérations suivantes sur chaque token pour:

1. Lemmatization
2. Étiquetage de partie du discours (POS tagging) en utilisant 7 catégories: (NP) nom propre, (N) nom, (A) adjectif, (V) verbe, (P) préposition, (D) déterminant, (O) other (autre).
3. Détection d’entités nommées (NER) en utilisant 4 catégories: (PER) Person, (ORG) Organization, (LOC) Location, (O) Other.

ATTENTION: Si vous préférez, utiliser la phrase française suivante: *Avec mes amis Ali et Yu, je suis allée au BlueBird Theater d’Ottawa, récemment rénové.* Rayez la première colonne, et continuer le traitement.

Tokenization	OU tokens français	Lemmatization	POS tagging	NER
I				
went				
to				
the				
movies				
at				
the				
renovated				
BlueBird				
Theater				
in				
Ottawa				
with				
my				
friends				
Ali				
and				
Yu				

PARTIE 6a - Représentation des connaissances (10 points)

Soit les mots suivants:

automobile, ballon, bicyclette, couleur, conduire, fruit, lancer, manger,
objet, orange, pomme, roue, rouge, route, table, véhicule

ATTENTION: Si nécessaire, vous pouvez introduire différents sens pour les mots pour répondre aux questions suivantes.

- a. Construisez une taxonomie avec le plus de noms possibles de la liste. Votre taxonomie peut être fragmentée, en ce sens qu'il est peut-être impossible de relier toutes les portions ensemble. N'ajoutez PAS de mots.

PARTIE 6a - Représentation des connaissances (10 points) (RÉPONSE continue)

- b. Construisez un réseau sémantique avec le sous-ensemble suivant de mots: bicyclette, roue, rouge, route, véhicule. Vous pouvez créer les relations sémantiques de votre choix. Votre réseau doit contenir au moins 4 liens et ne pas être fragmenté (cela doit être un seul réseau).

- c. Soit un cadre (*frame*) pour le verbe *lancer* contenant 3 rôles sémantiques (celui-qui-lance, objet-lancé, direction). Inventez deux phrases exemples pour illustrer les rôles sémantiques dans ce cadre.

PARTIE 7a - Architectures pour Apprentissage Profond (6 points)

Pour chaque énoncé, dites s'il est vrai (V) ou faux (F).

	Énoncé	V/F
1	Dans un auto-encodeur, le nombre d'unités cachées est toujours plus grand que le nombre d'unités de sortie.	F
2	Quand un auto-encodeur est utilisé pour le débruitage (denoising), du bruit est ajouté à la sortie avant l'entraînement.	F
3	Un auto-encodeur, et un réseau perceptron multi-couches peuvent être entraînés de la même façon.	V
4	Un réseau antagoniste génératif (<i>Generative Adversarial Network - GAN</i>) est utilisé pour réduire la dimensionnalité de l'entrée.	F
5	Dans un GAN, il y a deux réseaux qui apprennent à tour de rôle.	V
6	Dans un GAN, nous appliquons la fonction d'erreur (entropie croisée) à la sortie du réseau génératif.	F
7	Dans un réseau convolutif (CNN), la couche de max pooling apprend ses poids par descente de gradient.	F
8	Le pas (<i>stride</i>) dans un CNN aura un impact sur la taille de la carte d'activation résultante.	V
9	Dans une architecture typique d'un CNN, les couches seront empilées comme suit: couche de convolution, couche entièrement connectée, couche pooling, couche de convolution, couche entièrement connectée, couche pooling, pour autant de couches que désiré.	F

PARTIE 7b - Architectures pour Apprentissage Profond - CNN (9 points)

Étant donné cette image 5x5, et ce filtre 3x3 qui a pour but de détecter des lignes verticales dans des images.

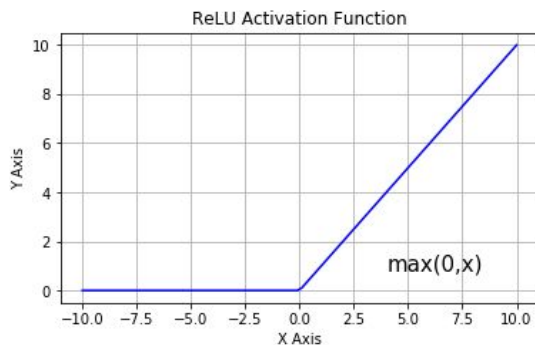
image					filter		
10	5	1	10	10	1	0	-1
10	5	1	10	10	1	0	-1
10	5	1	10	10	1	0	-1
10	10	10	10	10			
10	10	10	10	10			

Nous voulons faire une convolution du filtre sur l'image, en supposant un pas ("stride") de 1 et aucune marge ajoutée (aucun "zero-padding").

Montrez :

- le résultat (carte d'activation) de l'étape de convolution avant l'application de la non-linéarité.
- le résultat (carte d'activation) de l'application de la non-linéarité reLU sur la carte d'activation obtenue en (a).
- le résultat (carte d'activation) de l'application d'un max-pooling de taille 2x2 avec un pas de 1 sur la carte d'activation obtenue en (b).

Pour la question (b), le reLU est donné ci-bas.



PARTIE 7b - Architectures pour Apprentissage Profond (9 points) - RÉPONSE ICI

image

10	5	1	10	10
10	5	1	10	10
10	5	1	10	10
10	10	10	10	10
10	10	10	10	10

filter

1	0	-1
1	0	-1
1	0	-1

convolution

27	-15	-27
18	-10	-18
9	-5	-9

relu

27	0	0
18	0	0
9	0	0

Max pooling

27	0
18	0

PARTIE 8 - Projets (1 point)

Quel a été votre projet favori parmi toutes les présentations? Pas besoin de donner le titre exact, seulement le sujet.