



EXAMEN FOUILLE DE DONNÉE SUJET 3

EXAMEN DATA MINING

EXERCICE 1 : CLASSIFICATION PAR ARBRE DE DÉCISION 07PTS

Une société immobilière dispose des informations sur se clients :

Num	Emplacement	Type de maison	Revenu	client anterieur ?	Resultat
1	Banlieue	Unifamiliale	Élevé	Non	Insatisfait
2	Banlieue	Unifamiliale	Élevé	Oui	Insatisfait
3	Rural	Unifamiliale	Élevé	Non	Satisfait
4	Ville	Jumelée	Élevé	Non	Satisfait
5	Ville	Jumelée	Bas	Non	Satisfait
6	Ville	Jumelée	Bas	Oui	Insatisfait
7	Rural	Jumelée	Bas	Oui	Satisfait
8	Banlieue	Rangée	Élevé	Non	Insatisfait
9	Banlieue	Jumelée	Bas	Non	Satisfait
10	Ville	Rangée	Bas	Non	Satisfait

- 1 Définir le terme entropie et calculer celle de cette population.
- 2 Définir entropie résiduelle pour attribut A et calculer l'entropie résiduelle pour chaque attribut : Emplacement, Type de maison, Revenu, client antérieur.
- 3 Pour la construction de l'arbre de décision, utilisez-vous l'attribut revenu ? Pourquoi ?
- 4 Lors de la construction de l'arbre de décision. quel est l'attribut à tester à la racine de l'arbre.
- 5 Construire l'arbre de décision complet et élaguez le.
- 6 Quelle est le taux d'erreur de cet arbre estimé sur l'ensemble des clients 1 à 10.
- 7 Donner un intervalle de valeurs pour l'erreur réelle en utilisant une confiance de 90 %.

ON SE DONNE LES 4 CLIENTS SUIVANTS :

Num	Emplacement	Type de maison	Revenu	client anterieur ?	Resultat
11	Banlieue	Rangée	Bas	Oui	Satisfait
12	Rural	Rangée	Élevé	Oui	Satisfait
13	Rural	Unifamiliale	Bas	Non	Satisfait
14	Ville	Rangée	Élevé	Non	Insatisfait

8. Comment chacun de ces clients est-il classé avec l'arbre de décision que vous avez proposé dans l'item 5 ?



9. Pour ces 4 clients, on apprend par ailleurs que les clients 11 et 12 sont des clients réguliers, et que les clients 13 et 14 ne le sont pas. Quel est le taux d'erreur estimé sur les clients 11, 12, 13 et 14 ? Combien y a-t-il de faux positifs et de faux négatifs ?

EXERCICE 2 : RESEAU DE NEURONES 05PTS

- Définir des termes suivants : (a) Neurone, (b) Perceptron, (c) Fonction cout, (d) descente de gradient, (e) Convolution, (f) Couches, (g) Fonction d'activation ;
- Faire le schéma annoté d'un perceptron a 3 couches ;





클릭하기

3. Faire un schéma annoté d'un perceptron simple recevant des entrées x_1 , x_2 et x_3 d'un biais b et retournant une sortie y . Vous pourrez pondérer les synapses par des poids w





클릭하기

4. Pour ce perceptron de la question précédente, donnez l'équation de la fonction d'agrégation $z(x_1, x_2, x_3)$.

EXERCICE 3 : GENERATION DE RÈGLES D'ASSOCIATION 08PTS

Soit X la base de transactions contenant un ensemble de transactions décrivant des achats de produits dans le super marché FreeMarket.

Liste	Produits					
L1	M	O	N	K	E	Y
L2	D	O	N	K	E	Y
L3	M	A	K	E		

Liste	Produits					
L4	M	U	C	K	Y	
L5	C	O	O	K	I	E

1. Montrer que l'union de deux itemsets fréquents n'est pas toujours fréquent ;
2. A l'aide de l'algorithme Apriori, générer les règles d'association avec $\text{minConf} = 1$ pour les données de la base X ;
3. En considérant seulement l'ensemble de produits $\{M, O, N\}$, trouver l'ensemble de règles qui permettent de prédire l'achat de deux produits tout en améliorant la prédiction par rapport à la mesure statistique induite par la base de données ;
4. Répéter la même question pour trouver l'ensemble de règles qui permettent de prédire l'achat d'un produit.
5. Utilisez l'algorithme fp-growth pour générer la liste des itemsets fréquent de la base X avec $\text{minsup}=3$.









[클릭하기](#)

6. La réponse à la question 5 aurait été plus rapide (simple) en utilisant l'algorithme Apriori ? justifiez-vous.

스마트팜

농업회사법인 주식회사 그린(Griin)





CORRIGE EXAMEN FOUILLE DE DONNEE :

EXERCICE 2:

1) Les définitions :

- 1
 1. Neurome : Le terme correct est "neurone". Un neurone est une unité de base du système nerveux qui permet de recevoir, traiter et transmettre l'influx nerveux.
- 2
 2. Perceptron : Un perceptron n'est pas un algorithme d'apprentissage, mais plutôt un modèle de neurone artificiel qui peut être utilisé dans des problèmes de régression ou de classification. Il s'agit d'un modèle simplifié qui prend des entrées pondérées, les

agrège et applique une fonction d'activation pour produire une sortie.

3. Fonction coût : Une fonction coût est utilisée pour quantifier l'écart entre les prédictions d'un modèle de réseau de neurones et les valeurs réelles attendues. Elle mesure les erreurs du modèle pendant la phase d'apprentissage.
4. Descente de gradient : La descente de gradient est un algorithme d'optimisation utilisé pour ajuster les paramètres d'un modèle de réseau de neurones en calculant les gradients de la fonction coût par rapport à ces paramètres. L'objectif est de trouver les valeurs des paramètres qui minimisent la fonction coût, rendant ainsi le modèle le plus précis possible.
5. Convolution : L'opération de convolution est effectivement fondamentale dans les réseaux de neurones convolutifs (CNN). Elle consiste à appliquer un filtre (noyau) sur une image (ou une autre entrée) pour extraire des caractéristiques spécifiques de l'image en effectuant des opérations de multiplication et d'agrégation.
6. Couche ou layer : Une couche dans un réseau de neurones est une disposition verticale d'un ensemble de neurones qui reçoivent des entrées et produisent des sorties. Chaque neurone dans une couche est connecté aux neurones de la couche précédente ou suivante.
7. Fonction d'activation : Une fonction d'activation est une fonction mathématique appliquée à la sortie d'un neurone (ou d'une couche de neurones) pour introduire de la non-linéarité dans le modèle. Elle permet au modèle de capturer des relations complexes entre les entrées et les sorties.
8. Perceptron multicouche : Un perceptron multicouche (MLP) est un type de réseau de neurones artificiels composé de plusieurs couches de neurones, y compris une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Les neurones sont connectés de manière séquentielle, avec des connexions pondérées entre les couches.
9. Neurone artificiel : Un neurone artificiel est une abstraction mathématique inspirée des neurones biologiques. Il est utilisé pour résoudre des problèmes de logique binaire à la base, en prenant des entrées pondérées, en les sommant et en appliquant une fonction d'activation pour produire une sortie.
10. Neurone biologique : Un neurone biologique est une unité fondamentale du système nerveux des êtres vivants. Il permet de transmettre l'influx nerveux des zones d'excitation aux zones de traitement à travers l'axone.

2)

- 1 Un perceptron a 3 couches : (*couche d'entrée de 5 cellules, une couche cachée de 3 neurones et une couche de sortie de deux neurons*) :

- 1 Je vais pendre un exercice de mon site pour cette question (<https://www.pandacodeur.com/pages/examen-pandacodeur/examen-fouille-de-donnee/examen-fouille-de-donnee-sujet-2.html>) : Nous avons l'architecture la plus simple du réseau de neurones : un réseau avec une seule couche d'un seul neurone

Nous utilisons 3 données en entrée :

- > x1 - la pluviométrie de l'année,
- > x2 - la température moyenne sur l'année, et
- > x3 - le type de sol (on va considérer que chaque type de sol a été labellisé par un numéro).

- 1 Une **fonction d'agrégation**, qui permet de calculer une unique valeur à partir des entrées et des poids correspondants.

Exemple en image :

En fonction de l'exercice de la question (3) La combinaison linéaire des données d'entrée (x1, x2, x3) pondérées par les poids correspondants (w1, w2, w3) et additionnées avec le biais (b) est exprimée comme suit :