

Rédigé par : Astrid Jourdan, Jean-Paul Forest, Hervé de Milleville

Ref :

A l'intention de : Elèves d'ING2-GI

Créé le : 02/02/2016

## 1. Préambule

Cet examen dure 1h30. Il porte sur les acquis théoriques de la matière.

Aucun document, ordinateur, calculatrice, téléphone portable n'est autorisé.

## 2. Identification de tâches

Le CROUS (CLOUS) est un organisme de services aux étudiants notamment concernant la restauration. Sur un campus se trouvent généralement plusieurs restaurants universitaires. Chaque restaurant propose ses propres plats, viande, poisson, pizza, frites, pâtes,...

La fréquentation des restaurants dépend de certaines périodes (congs, examens, jour de la semaine,...).

Chaque étudiant est pourvu d'une carte nominative. Lors du passage en caisse la carte est lue, donc on peut connaître l'identité du consommateur (âge, type de formation, boursier,...), le montant, la date,....

Identifier trois tâches différentes en précisant qui sont les instances (individus), les attributs, l'objectif de la tâche, et s'il s'agit d'une méthode supervisée ou non et pourquoi.

## 3. Clustering

Dans IR, on considère 5 points  $x_1=1$ ,  $x_2=2$ ,  $x_3=9$ ,  $x_4=12$  et  $x_5=20$ .

3.1. Appliquer l'algorithme des k-means avec les valeurs de k et les points de départ suivants.

- $k=2$ ,  $c_1=1$  et  $c_2=20$
- $k=3$ ,  $c_1=1$ ,  $c_2=9$  et  $c_3=12$
- Peut-on utiliser l'inertie intra-classes (ou inter-classes) pour déterminer le meilleur regroupement des deux ? Justifier votre réponse.

3.2. Appliquer une méthode de classification hiérarchique ascendante en utilisant la distance minimale comme critère de dissimilarité entre classes. Tracer le dendrogramme (avec en ordonnée la distance minimale). Quel regroupement vous paraît correct ?

## 4. Méthodes

4.1. Laquelle/lesquelles de ces méthodes (algorithmes) peut-on utiliser en apprentissage non supervisé ?

- Recherches d'associations 
- k-means
- Classification hiérarchique
- Plus-proches voisins 
- Arbres de décision 

4.2. Parmi les algorithmes vus en cours lesquels sont stochastiques (ont un résultat aléatoire). Expliquez pourquoi.

4.3. Quelles sont les différences entre un modèle ayant de bonnes performances sur son ensemble d'apprentissage et un modèle ayant de bonnes performances sur un ensemble test ?

4.4. Que mesure l'entropie ?