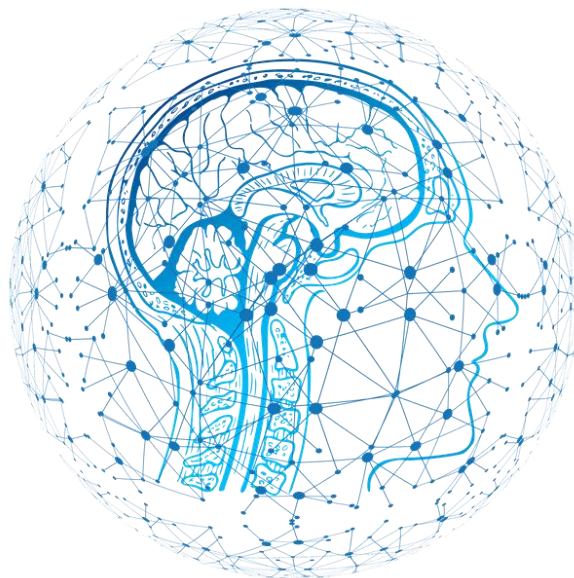

INTELLIGENCE ARTIFICIELLE

Projet Auto-encodeur et classification



12 MAI 2023

Alexandre DUBERT et Hayk ZARIKIAN

1 Introduction

Ce projet est un projet proposé par l'UE « Intelligence artificielle » de l'Université de Strasbourg – UFR Mathématique et Informatique. Nous avons décidé de travailler sur ce projet en binôme afin qu'on puisse se questionner sur les différentes questions concernant l'apprentissage et la classification des données par l'IA.

2 Données

2.1 Question 2.1

Les prix sont encodés sous forme de classe qui correspond :

Classe 0 : low cost

Classe 1 : medium cost

Classe 2 : high cost

Classe 3 : very high

2.2 Question 2.2

Le paramètre « d » représente les attributs dans le dataset ce qui correspond à d= 20 attributs.

2.3 Question 2.3

Le nombre de classe représenté dans le dataset correspond à k = 4

2.4 Question 2.4

Nombre d'exemple dans chacune des classes sont les suivants :

Pour les données d'entraînement :

Classe 0 : 302 exemples

Classe 1 : 298 exemples

Classe 2 : 304 exemples

Classe 3 : 296 exemples

Pour les données de test :

Classe 0 : 198 exemples

Classe 1 : 202 exemples

Classe 2 : 196 exemples

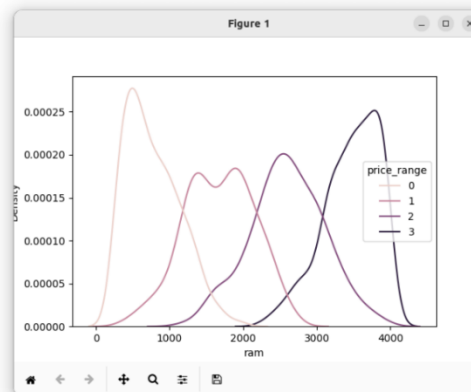
Classe 3 : 204 exemples

2.5 Question 2.5

Consulter le fichier « projet.py »

2.6 Question 2.6

Le facteur le plus discriminant quant au prix est la RAM. En effet, sur la capture, nous constatons une réelle et nette différence entre les différentes catégories.



2.7 Question 2.7

État d'avancement : Fonctionnel

Consulter le fichier « projet.py »

3 Évaluation

3.1 Question 3.1

Évaluer le modèle sur un jeu de données différent permet de s'assurer qu'il est capable d'extrapoler ses connaissances sur les données qui n'ont pas servi à l'apprentissage.

3.2 Question 3.2

Lorsque EP vaut 1 toutes les données ont été correctement classées donc $T1 = T2 = T3 = T4 = 1$

3.3 Question 3.3

A l'inverse lorsque EP vaut 0 aucune donnée n'a été correctement classée donc $T1 = T2 = T3 = T4 = 0$

4 Algorithmes

4.1 Apprentissage de l'auto-encodeur

4.1.1 Question 4.1.1

L'entropie croisée n'est pas adaptée comme critère d'optimisation dans les situations où la sensibilité aux erreurs de classification des différentes classes est différente. Elle peut être utilisée lorsque la distribution des classes est équilibrée ou lorsque l'on souhaite simplement minimiser l'erreur globale sans tenir compte de la sensibilité aux erreurs de chaque classe. Dans les cas où la sensibilité aux erreurs de classification est différente, il est préférable d'utiliser des critères d'optimisation tels que le taux d'erreur par classe, la précision par classe.

4.1.2 Question 4.1.2

La fonction d'activation de la dernière couche et la fonction tangente hyperbolique qui renvoie des valeurs entre -1 et 1, donc pour pouvoir les comparer à l'entrée du réseau, il faut que l'entrée soit normalisée.

4.1.3 Question 4.1.3

État d'avancement : Fonctionnel

Consulter le fichier « neurones/utility.py »

4.1.4 Question 4.1.4

L'apprentissage de l'auto-encodeur n'est pas supervisé, car il ne nécessite pas d'étiquettes de classe pour entraîner le modèle. Il utilise simplement les données d'entrée pour reconstruire les données de sortie correspondantes, sans avoir besoin d'informations supplémentaires sur la classe ou la catégorie des données.

4.1.5 Question 4.1.5

État d'avancement : Fonctionnel

Consulter les fichiers « neurones/reseau.py »

4.2 Apprentissage de l'arbre de décision dans l'espace latent

4.2.1 Question 4.2.1

État d'avancement : Partiellement

Consulter le fichier « neurones/reseau.py »

4.2.2 Question 4.2.2

État d'avancement : Partiellement

Consulter le fichier « arbre/arbre.py ». Nous avons construit un arbre mais il n'arrive pas à prédire les classes de prix à partir des données encodées

4.2.3 Question 4.2.3

Les données encodées sont de dimension 2 donc il y aura beaucoup moins de classes discriminantes que sur le jeu de donnée initial (20 attributs). Donc l'arbre sera moins profond.

4.2.4 Question 4.2.4

État d'avancement : Non fait

5 Conclusion

Pour conclure, ce projet nous a beaucoup apporté dans la compréhension du fonctionnement de la classification et l'apprentissage de données via l'intelligence artificielle car les questions théoriques nous font réfléchir sur le fonctionnement d'un auto encodeur et de la classification des données. Par ailleurs, les questions pratiques nous montrent comment il est possible d'exploiter ce genre de donnée dans un programme qui permet de tenir des conclusions et des comparaisons.