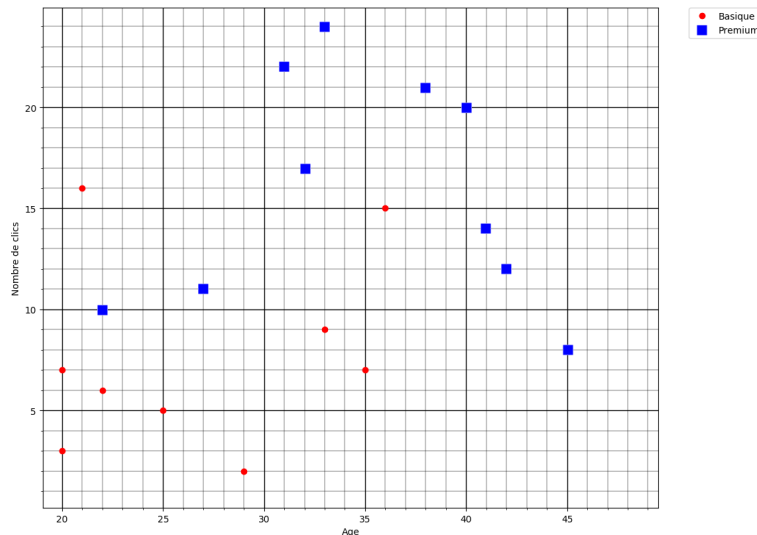


L'algorithme des k plus proches voisins (ou kNN :)

1 Introduction : Problème de classification

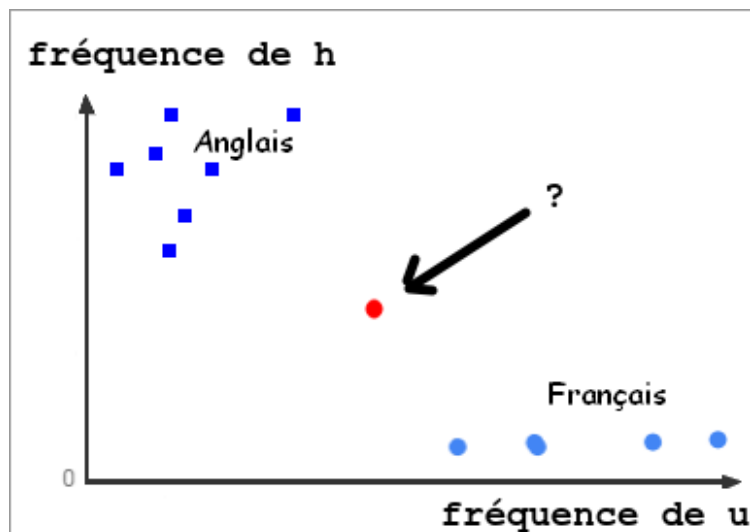
Exercice n°1

Une société "Amazing" souhaite faire de la publicités ciblée selon les caractéristiques des clients. Voici les données recueillies sur quelques clients :



1. Bob, 43 ans nouvel utilisateur du site effectue en moyenne 18 clics par jour.
 - En observant son plus proche voisin, peut-on proposer à Bob un abonnement Premium ?
 - En observant ses 3 plus proches voisins, peut-on proposer à Bob un abonnement Premium ?
2. Bob Junior, 23 ans nouvel utilisateur du site effectue en moyenne 9 clics par jour.
 - En observant son plus proche voisin, peut-on proposer à Bob Junior un abonnement Premium ?
 - En observant ses 3 plus proches voisins, peut-on proposer à Bob Junior un abonnement Premium ?

Exercice n°2



Remarques :

- Pour construire de tels graphiques on a besoin
- Comment calculer une "distance" ?

Exercice n° 3

A l'aide du fichier python, compléter les tableaux ci-dessous au fur à mesure.

Texte	freq u	freq h	étiquette
texte0.txt			Anglais
texte1.txt			Anglais
texte2.txt			Français
texte3.txt			Français
texte4.txt			Anglais
texte5.txt			Français
texte6.txt			Français
texte7.txt			Anglais
texte8.txt			Anglais
texte9.txt			Français

Texte	distance du texte11.txt
texte0.txt	0.0249
texte1.txt	
texte2.txt	
texte3.txt	
texte4.txt	
texte5.txt	
texte6.txt	
texte7.txt	
texte8.txt	
texte9.txt	

En déduire la langue du fichier texte11.txt :

2 Apprentissage

Trois types d'apprentissage :

Qu'est ce qui fait qu'un apprentissage est bon ?

Exercice n° 4 Lire l'article ci-dessous et surligner les passages en lien avec le chapitre.

Tinder et Elyze: l'algorithme est-il déterministe?
par Aurélie Jean

Qui n'a pas entendu parler d'Elyze ? Présentée comme le Tinder du candidat idéal, cette application mobile vous suggère le candidat à la présidentielle pour lequel voter d'après vos réponses au quiz censé capturer vos convictions politiques. Alors qu'on aimerait croire au partenaire et au candidat parfaits sur un simple *swipe* à droite, les choses ne sont pas si simples. Les algorithmes de *matching*, qui font correspondre des cœurs ou des intentions de vote, objectivent avec une confiance déconcertante l'amour ou la sensibilité politique. Mais les comprendre nous évitera de tomber dans une passion aveugle et nous permettra de bien maîtriser ces applications.

Tinder et Elyze sont des applications dites de « *matching* », qui vous suggèrent donc un ou des individus idéaux, selon l'objectif recherché, sentimental ou politique. En pratique, ces applications utilisent des données du profil statique et dynamique des utilisateurs pour réaliser

cette suggestion, en théorie ultrapersonnalisée, qui fait coïncider des personnes d'après un calcul algorithmique. Le profil statique est lié à des informations que l'utilisateur entre explicitement au moment de l'enregistrement, comme le genre, l'âge ou encore la localisation, alors que le profil dynamique correspond au comportement de l'utilisateur : d'autres profils ou des idées d'un programme politique que l'utilisation semble apprécier (le fameux *swipe* à droite). Mais que valent réellement ces suggestions prédictives d'un certain avenir ?

Tout d'abord, il faut parler de suggestion et non de réponse absolue. Nous pouvons à tout moment rejeter la proposition que la machine nous offre. Aussi, les algorithmes de *matching* ne sont pas des algorithmes dits « déterministes », et ce pour plusieurs raisons. Un algorithme déterministe est un algorithme qui, pour un jeu de données d'entrée, donnera toujours le même résultat, peu importe le moment où il est exécuté. Dans le cas d'un algorithme de *matching*, il peut y avoir des variables pseudo-aléatoires, ou tout simplement un apprentissage au fur et à mesure du temps qui fait évoluer les réponses. Enfin, il ne faut pas écarter les phénomènes d'autoréalisation – vous votez pour une personne, car on vous l'a suggéré – et de biais de confirmation – vous connaissez des couples mariés qui se sont rencontrés en ligne – qui déforment notre perception de l'efficacité de l'outil en question.

Même si l'application parfaite existait, il ne serait pas recommandé de l'utiliser de manière unilatérale, mais plutôt pour éclairer notre jugement. En effet, en amour, les relations passées ont pour beaucoup d'entre elles eu raison de leur existence, car elles nous ont fait grandir, nous ont permis de mieux nous connaître, ou nous ont permis d'apprendre, tout simplement. De la même manière, en politique, nos choix, qui changent avec le temps, nos discussions parfois houleuses et nos propres remises en question nous permettent d'approfondir notre réflexion sur le candidat finalement retenu, et d'entraîner notre citoyenneté dans sa fonction majeure : voter.

Rien n'est déterministe, tout est mouvant et parfois incapturable... et c'est le philosophe Pascal qui le dit sur l'amour ■

FIGURE 1 – Le Point - 24 février 2022-