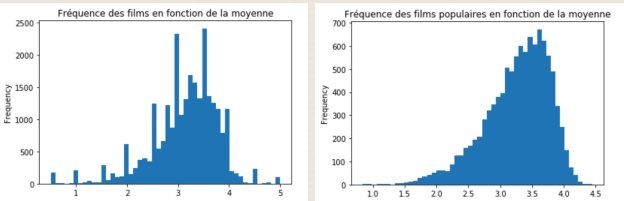


PROBLÉMATIQUE 1

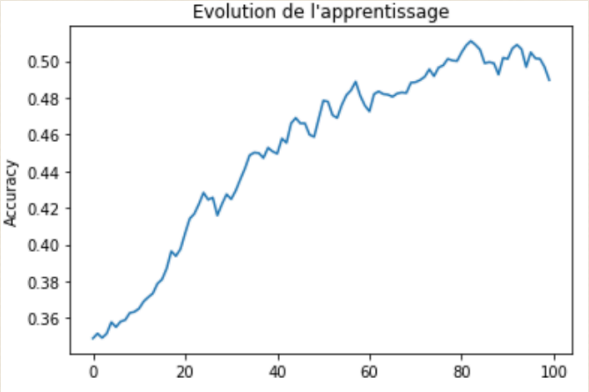
Description: Peut-on prédire la note d'un film en fonction de si il appartient à une catégorie ou pas ?
On filtre alors la base de donnée MovieLens en regardant seulement les films ayant eu plus de 100 avis pour ne pas avoir un résultat biaisé par des films ayant reçu seulement quelques avis et qui ne représentent donc pas une population homogène.



Avant (à gauche) la suppression des données non utiles et après

Outils d'analyse : Régression supervisé avec l'algorithme Moindre Carré

Résultat :



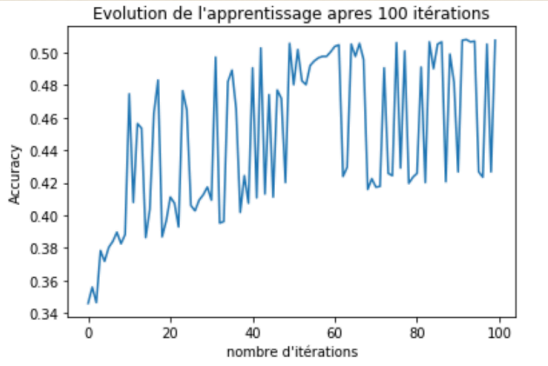
Au bout de 100 itérations on arrive à une accuracy de 0.4896694214876033

PROBLÉMATIQUE 2

Description : Peut-on trouver classer un film dans son siècle en fonction de sa note, de sa catégorie (ici on regardera en fonction de si le film est un Western ou un IMAX) et du nombre d'acteur qui joue dans ce film.

Outils d'analyse : Classification supervisée avec l'algorithme du gradient non stochastique

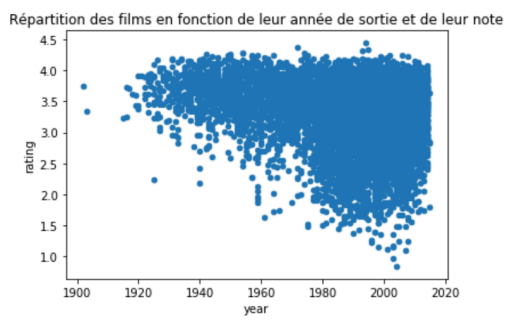
Résultat :



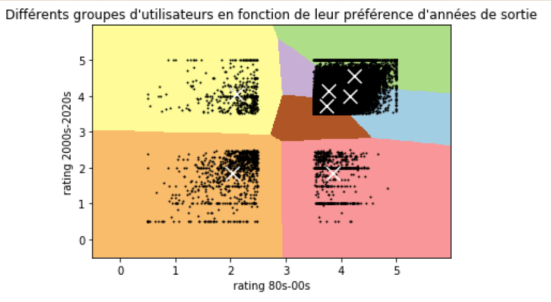
Au bout de 200 itérations on arrivé à une accuracy de 0.431924882629108

PROBLÉMATIQUE 3

Problématique : Peut-on regrouper des utilisateurs en fonction de leur préférence d'un film de la fin du 20ème siècle et du début du 21ème siècle ?
On peut remarquer que plus on avance dans le temps, plus des films ont des notes basses.



Outils d'analyse : Catégorisation non supervisée avec l'algorithme des k-moyennes



On peut alors voir alors 7 groupes différents d'utilisateurs. Ceux qui n'aiment pas trop les deux siècles, ceux qui aiment seulement un siècle, et ceux qui aiment les deux siècles. Dans ce dernier groupe on peut encore voir apparaître 4 groupes différents en fonction de plus ou moins leur note.