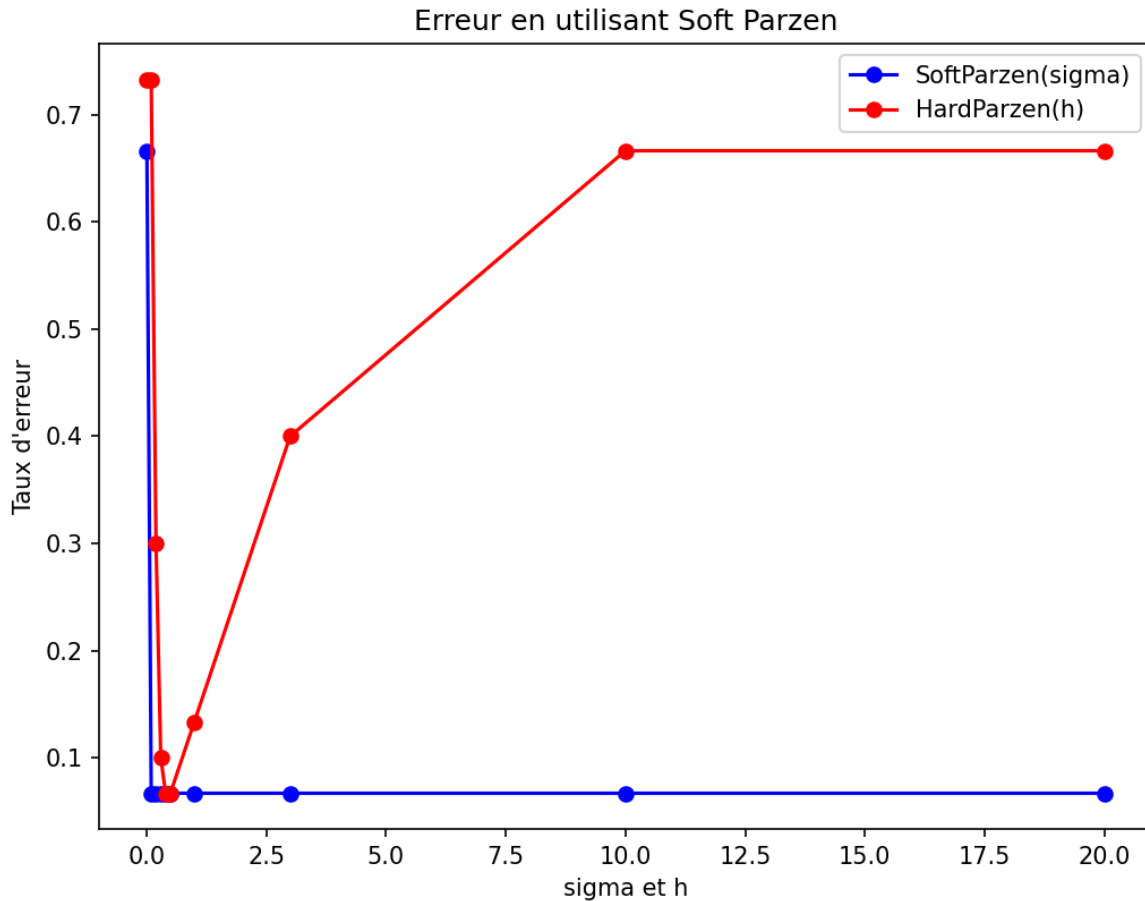


Graphe d'erreur



Discussion:

Discussion Q5: Dans le cas de Hard Parzen, il est évident qu'un paramètre h très petit est problématique, car il oblige à sélectionner les voisins de manière quasi aléatoire. Le taux d'erreur s'approche alors de 66 %, correspondant à une probabilité de 1/3 d'être correct. De plus, lorsque h est proche de 0,05, les points de test n'ont que très peu de voisins. En revanche, avec une fenêtre plus large, le modèle souffre de underfitting, car il inclut trop de points éloignés dans son estimation.

Discussion Q7 : La complexité de calcul du modèle Soft Parzen est indépendante du paramètre σ . En effet, on doit calculer la distance entre chaque point du jeu de test et tous les points du jeu d'entraînement, quelle que soit la valeur de σ . En revanche, pour Hard Parzen, plus le paramètre h est grand, plus la fenêtre englobe de voisins, ce qui augmente le nombre de distances à calculer. Ainsi, la complexité de Hard Parzen dépend directement de la taille de h .