Pour un prior uniforme ($P(\omega_i)=1/C$), les deux discriminadeurs donne les même performances. $P(\omega_i I \omega_j)= \omega_{i,j}/\sum_i \omega i, j$. Quand on change le choix de la fonction côut, enfait on a élargi l'impact négatif de $\omega_{1,2}$, donc dans la matrice de confusion, $\omega_{1,2}$ devient 0, mais en même temps, la valeur de $\omega_{2,1}$ a augmenté énormément, cela entraîne une baisse des performances du discriminateur. Quand on augmente la proportion de la deuxième classe, on a beaucoup moins de vecteur caractéristique dans la classe 1 et 3, donc à ce moment là, le risque de mal classifié est considérablement augmenté, le discriminateur donne une performance encore pire.

Quand on considère le problème de 10 chiffres, le discriminateur de Bayes linéaire et le discriminateur de Bayes ont des performances similaires, et ils ont des meilleurs performances comparé avec le discriminateur de Bayes quadratique. Le risque du discriminateur de Bayes est le plus bas parmi les trois discriminateurs ($<10^{-1}$), et les deux autres donne les même risques quand P_{app} est petit, mais quand P_{app} augmente, le discriminateur de Bayes linéaire est le premier à diminuer. L'utilisation de la deuxième fonction de côut conduit à la difficulté de distinguer le chiffre 1 et 8, et le choix du premier prior apporte des meilleur performances.

Question :pourquoi le risque du discriminateur de Bayes est si bas ?

La performance de RN_β pour le chiffre 8 est pire que RN_k et RN, mais les performances de RN_β et RN_k sont similaires. Pour ne pas avoir besoin d'effectuer un nouvel apprentissage à chaque fois, on peut geler la matrice W une fois on finit le processus d'entraînement.