SY09 Printemps 2014TP 3Théorie de la décision

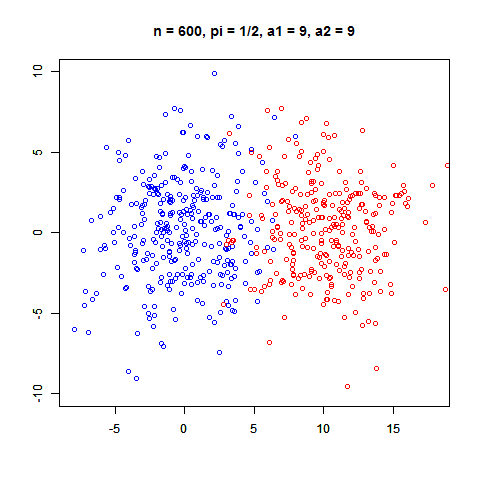
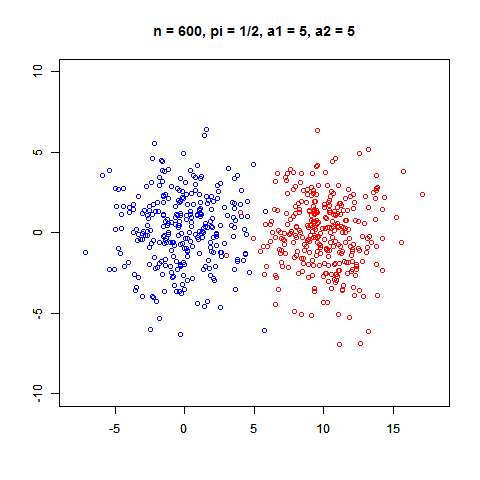
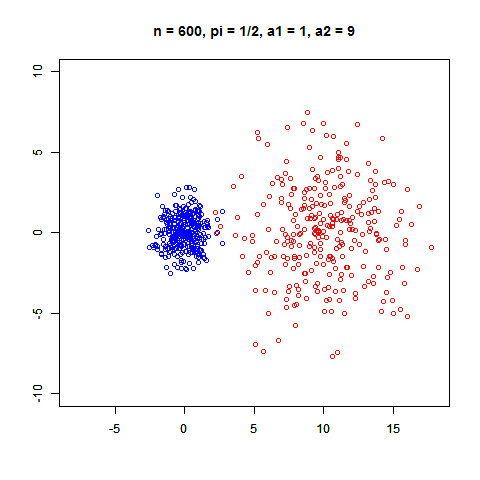
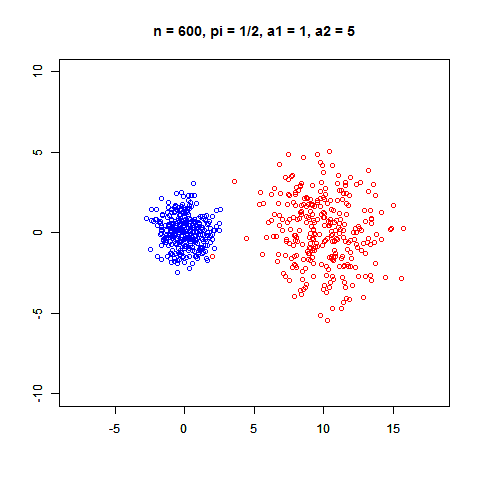
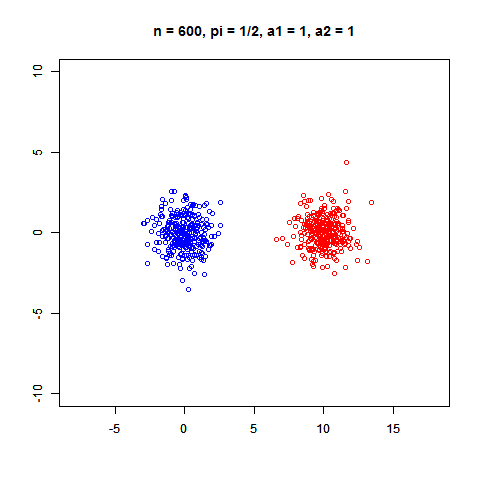
L’objectif de ce TP est de mettre en pratique deux théories de la décissions qui sont le classificateur euclidien d’une part ,nous allons alors tenter d’estimer son efficacité et la règles de Bayes d’autre part en étudiant un problème de reconnaissance de cible.

Exercice 1. Classifieur euclidien

Le but de cet execice est d’étudier les performances du classificateur euclidien sur des échantillons issus des deux classes ω1 et ω2 appartenant à R2  dont les distributions sont normales et de paramètres respectifs (μ1,Σ1) et (μ2,Σ2

* Simulation d’un échantillon :

En utilisant la fonction mvrnorm de la bibliothèque MASS, nous avons écrit la fonction simul qui nous retourne un échantillon de taille n tiré suivant une proportion π d'exemples issus de la classe ω1. Pour chaque exemple, nous avons, dans un premier temps, tiré au hasard la classe dont il est issu, avant de le générer en utilisant les paramètres adéquats :

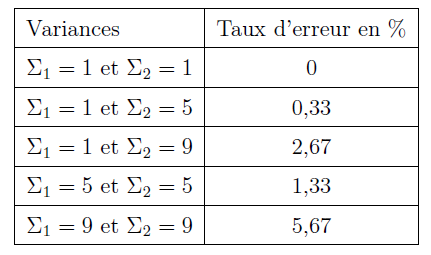


Grâce à ces simulations, nous observons que plus la variance est élevée plus la dispersion des point est grande (autour de la moyenne), ils sont moins concentrés. Il apparaît qu’à partir d’une certaine valeur de variance ‘exemple de la cinquième figure) il existe un chevauchement entre les classes ω1 et ω2. Il devient alors plus difficile de les distinguer et on peut d’ores et déjà supposer que la probabilité d’erreur sera plus importante.

Estimation de la probabilité d’erreur

Désormais nous allons chercher à estimer la probabilité d’erreur associée au classifieur Euclidien. Pour cela nous avons séparé en deux parties de même cardinal nos échantillons initiaux contenant les classes ω1 et ω2. La première partie est l’échantillon d’apprentissage permettant d’apprendre les moyennes μ1 et μ2. La seconde partie est l’ensemble de tests permettant d’estimer le taux d’erreur de classifieur Euclidien sur ces données.

Utiliser un ensemble de test nous permet d’estimer la performance du classifieur. Nous avons pour celà tout d’abord écris une fonction règle Euclidienne qui correspond au classifieur Euclidien. Cette fonction compare les distances entre d’une observation x et l’estimateur μ1 puis μ2 et renvoie alors la classe dont x est le plus proche. Ensuite nous avons réalisé une fonction erreurEstimee qui calcul le nombre d’erreur de classification que notre classifieur a commis sur l’ensemble de test divisé par le cardinal de cet ensemble. Cette fonction renvoie donc le pourcentage d’erreur estimé.

Voici les résultats concernant nos 5 simulations 

Les résultats ci-dessus nous montrent bien que plus la variance est élevée, plus le taux d’erreur augmente et donc moins le classifieur Euclidien est performant. Ceci s’explique par le fait que la règle Euclidienne évalue la distance euclidenne entre chacun des points et les centres des deux classes. Il attribue ensuite à ce point la classe dont le centre est le plus proche.

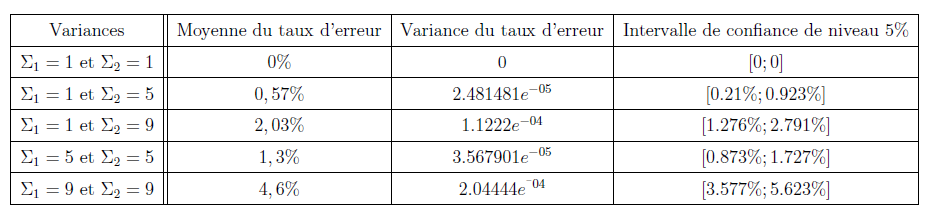
De ce fait, lorsque la dispersion d’une classe est grande, les échantillons ont tendance à se chevaucher. Ainsi certains points se retrouve mal classifiés car il peut se retrouver plus proche du centre de l’autre classe si c’est un point éloigné du centre de sa propre classe (grande dispersion), le taux d’erreur est alors plus élevé. En ce qui concerne la simulation 4, où Σ1 = 5 et Σ2 = 5,

Nous remarquons que le taux d’erreur est moins élevé que la simulation 3 où Σ1= 1

et Σ2 = 9. Cela laisse à penser qu’en terme d’efficacité il vaut mieux disposer de deux classes de variances modérées que d’une classe de variance faible et une autre très élevée. En effet bien que les deux sommes des variances de chaque simulation soient égales (10) une variance très élevée rendra un taux d’erreur plus important car se rapprochera plus facilement du centre de la classe à variance réduite (il n’y aura qu’un seul type d’erreur).

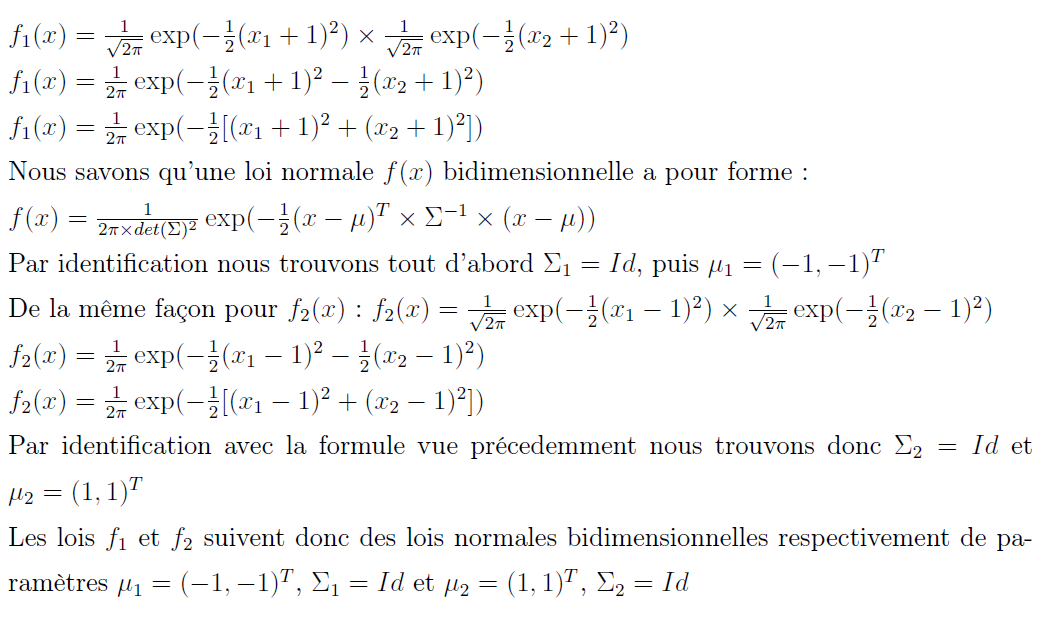
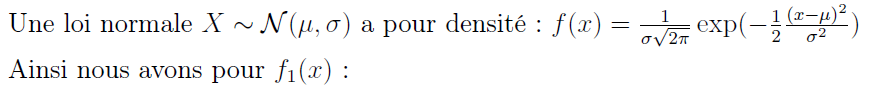
Probabilité d’erreur moyenne

Nous cherchons maintenant a observer la probabilité d’erreur moyenne, nous avons créé une fonction permettant de répéter les opérations précédentes 10 fois. À partir de ces dix simulations pour chacun des cas nous avons calculé la moyenne, la variance et un intervalle de confiance de niveau 5% sur l’espérance de la probabilité d’erreur. L’intervalle de confiance étant calculé grâce au test de Student.

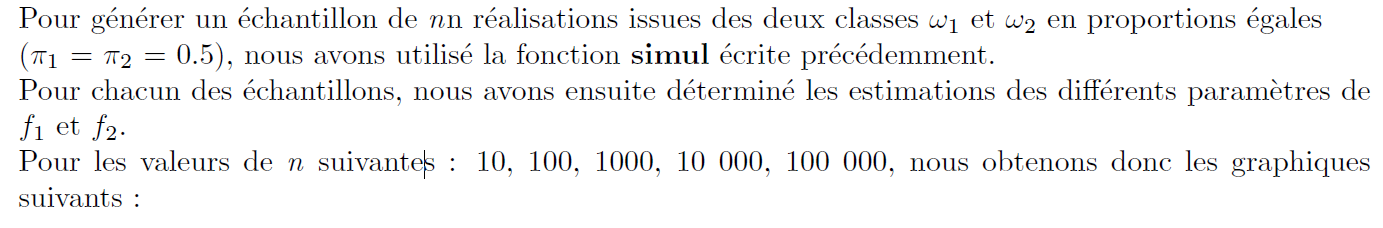


Nous remarquons que répéter l’opération permet d’abaisser le pourcentage d’erreur et de rendre la classifieur euclidien plus efficace.

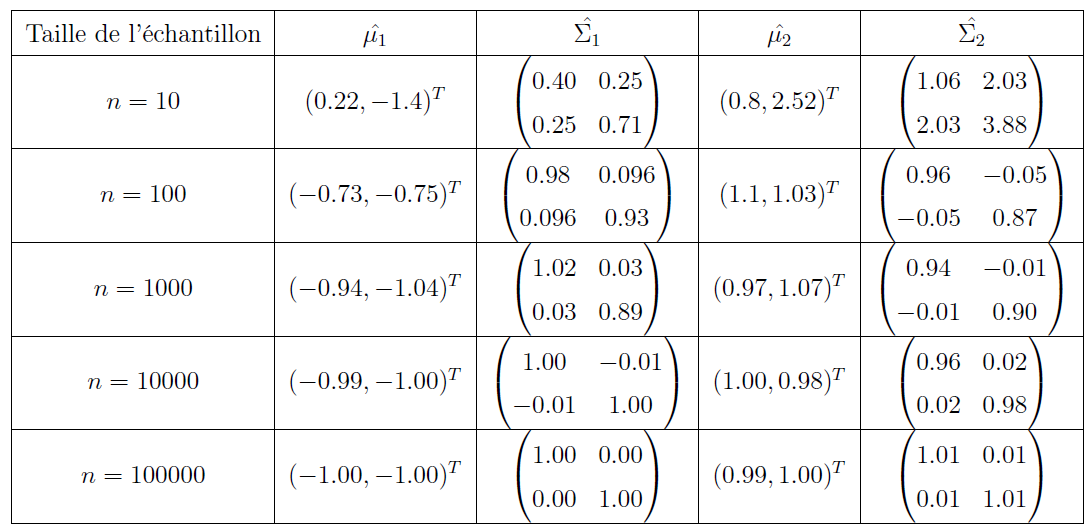
Exercice 2. Règle de Bayes

**1)**

**2)**

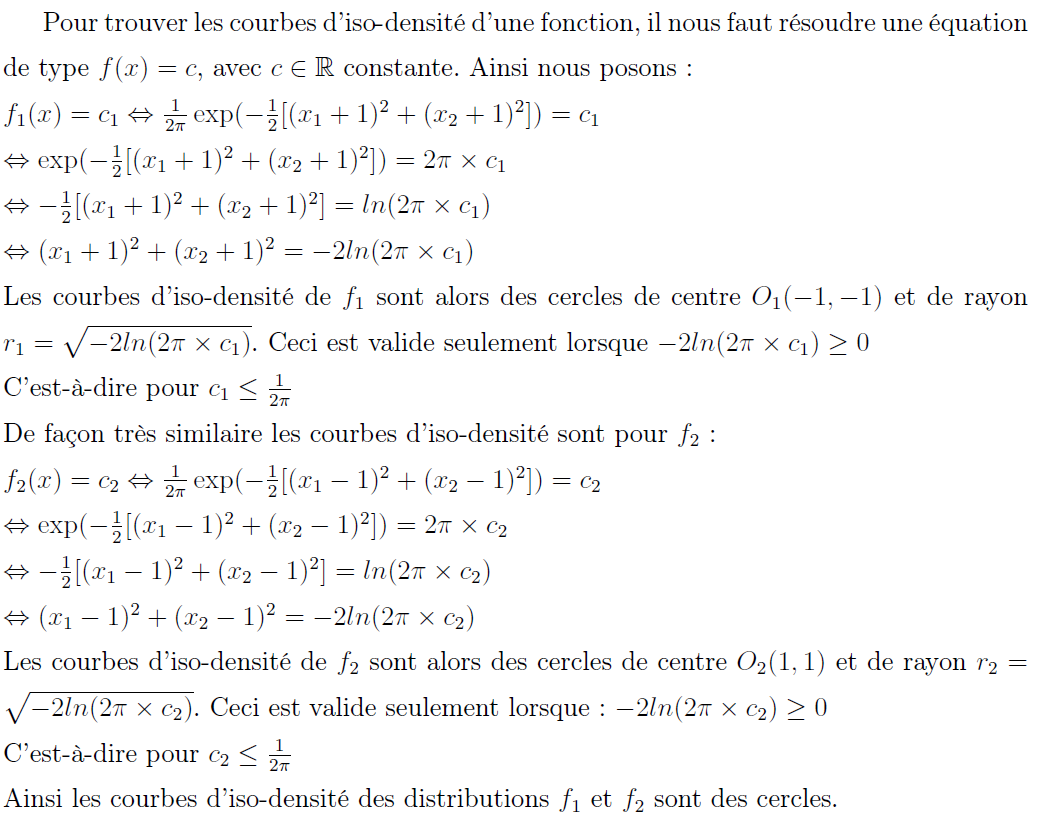


Résultats :



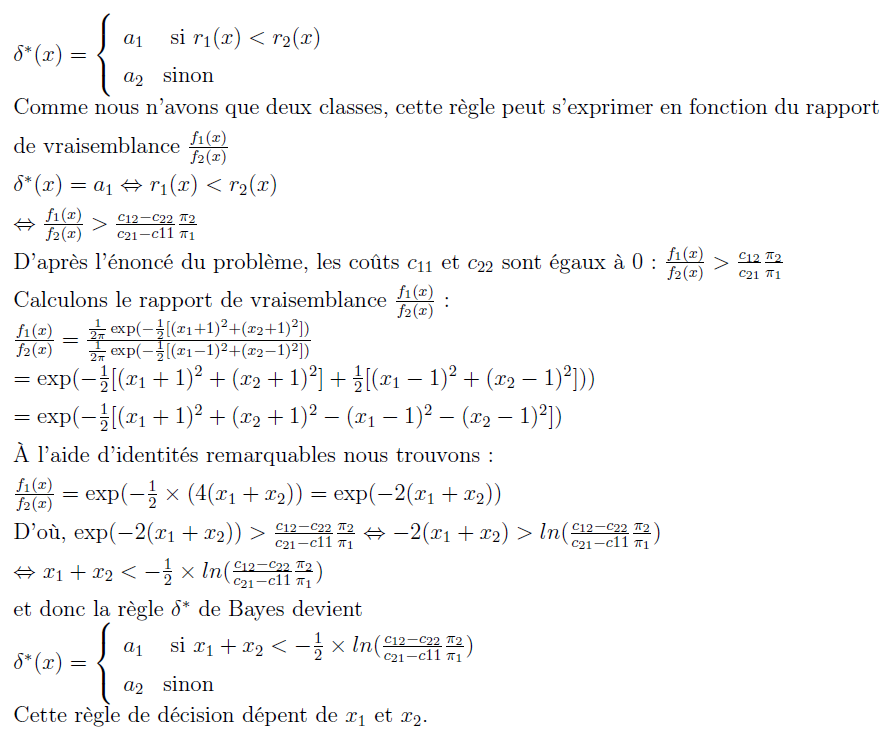
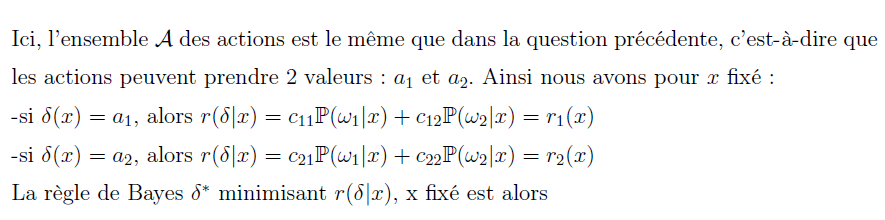
Nous avons pu observer que plus n était grand, plus on se rapprochait de la valeur théorique.

**3) Courbes iso densité**

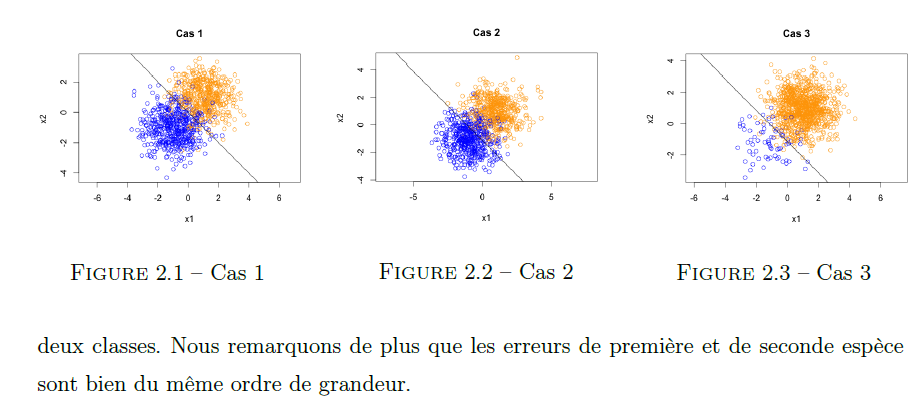
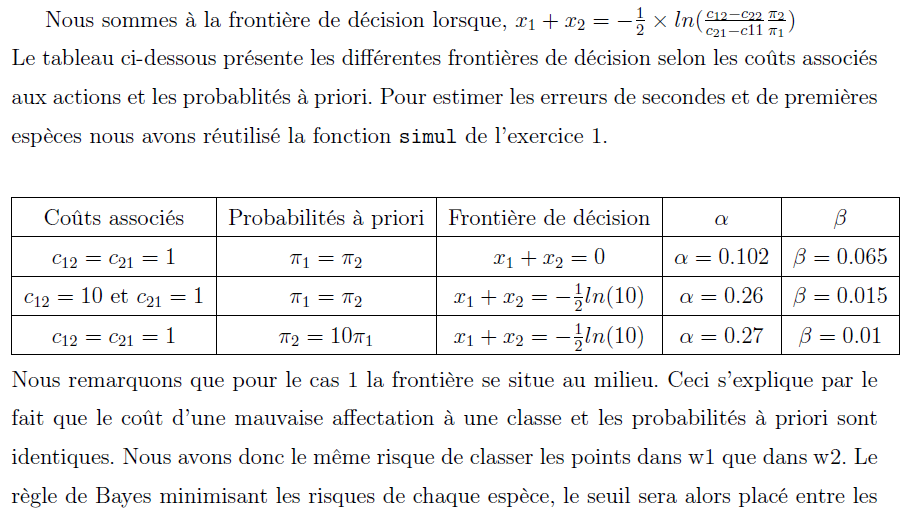


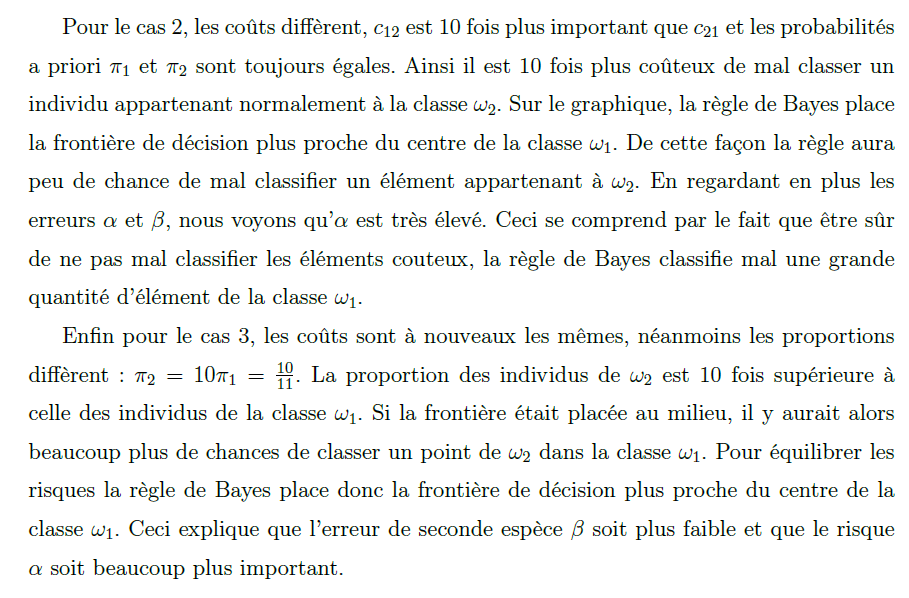
**4) Règle de Baynes**

a)



B + C)





Bon trop la flemme je vois mm pas trop quoi modifier exactement juste des petite tournure…

Et pour le dernier l’avant dernier paragraphe (ligne 6) il manque un « pour » je crois : « ceci se comprend par le fair que « pou » être sure… »

CONCLUSION

Nous avons finalement pu dans ce TP étudier un classifieur euclidien et les effets des valeurs des variances sur le pourcentage d’erreur. En outre nous avons pu observer que comme dans beaucoup de cas, répéter l’expérience permettait de réduire ce pourcentage d’erreur.

Dans le deuxième exercice nous avons pu étudier la règle de Bayes et observer l’influence de la proportion des classes et des coûts sur les frontières de décisions, la rêgles de Bayes tentant de minimiser le pourcentage d’erreur et le coût global.

Les règles de décisions permettent donc de classifier des éléments en fonction d’observations et de paramètres prédéfini, il est alors possible de jouer sur ces paramètres et sur la taille de l’échantillon afin d’influencer ces décisions.