Lilien (利廉) – 76246003
Magalie (周诗瞳) – 16241018

Compte rendu de TP (risque de Bayes)

1) Les figures 1 et 3 sont respectivement la répartition des classes estimées à partir du discriminateur « μ et Γ connus » et du discriminateur de « Bayes ». Les figures 2 et 4 sont le tableau croisé des classes vraies et classes estimées correspondant aux répartitions respectives des deux discriminateurs. Les trois diagrammes du haut de la figure 13 représentent la probabilité de x dans chacune des trois classes tandis que les trois diagrammes du bas représentent les zones des trois classes, ce qui nous offre une meilleure visualisation de la répartition des classes. A partir de la matrice de confusion, on peut estimer la probabilité d'estimer $P(\omega_i \mid \omega_j)$ en effectuant le rapport du nombre d'individus dans une case avec la somme des trois cases de la colonne. On remarque que les taux de généralisation et risques estimés sont égaux car les deux discriminateurs possèdent le même critère. En effet, le prior de Bayes est équiréparti donc il n'influence pas le discriminateur de Bayes. Ainsi, les matrices de confusion sont identiques. En modifiant le choix de la fonction coût, on remarque tout d'abord que le discriminateur de μ et Γ connus reste inchangé ce qui est logique puisque α n'intervient pas dans le calcul de ce discriminateur. En revanche, cette nouvelle condition contribue à agrandir la zone de la classe 2 au détriment de la classe 1, ce qui diminue l'erreur d'estimation de la classe 2 et augmente l'erreur d'estimation de la classe 1. Enfin, en augmentant le nombre d'individus de la classe 2, le discriminateur de Bayes arrive à être plus performant : il arrive à répartir la quasi-totalité des individus dans leur classe respective. Cela est dû au fait que l'on modifie le prior du discriminateur de Bayes, ce qui modifie son critère de discrimination.

2) Les coefficients de corrélation représentent le degré de similitude entre les deux nombres. On l'obtient en calculant le produit scalaire des vecteurs correspondant aux contours de deux nombres. On remarque que plus les coefficients de la matrice de confusion sont élevés, plus le carré correspondant est noir. Le risque du discriminateur de Bayes est indépendant de la taille de la base d'apprentissage tandis que le risque du discriminateur linéaire et quadratique diminue lorsque la taille de la base d'apprentissage augmente. Ainsi, dans chaque expérience, on doit seulement changer α et $P(\omega_j)$ sans avoir besoin de réapprendre. Le risque tend alors à converger vers le risque du discriminateur de Bayes classique. On peut voir que le risque de Bayes linéaire est inférieur au risque de Bayes quadratique. Cela est dû au fait que Bayes linéaire a seulement une matrice covariance tandis que Bayes quadratique en a 10. Ainsi, le Bayes quadratique a besoin d'une grande base d'apprentissage. Il est possible de recalculer le risque sur une nouvelle base de généralisation sans avoir besoin de relancer l'apprentissage car $P(x \mid \omega_j)$ reste fixé pour la généralisation. On constate bien les modifications attendues lorsqu'on modifie la base de coût et/ou la base de Prior sur la base de généralisation puisque la base d'apprentissage n'a pas été modifiée. Il n'y a donc pas de difficulté à utiliser la théorie du risque Bayésien.

3) Dans le cas de β, on pondère chaque entrée η_i par une valeur comprise entre 0 et 1 ramenée à la somme de tous les β_i. Dans le cas de κ, on pondère chaque entrée η_i par une valeur variant de 0 à l'infini ramenée à la somme de tous les κ_i. Cela permet d'éviter le calcul de la couche neuronale de sortie, on utilise alors directement les η_i comme sortie. Par comparaison du risque, le discriminateur RNβ est plus performant que le discriminateur RNκ, lui-même plus performant que le discriminateur RN classique. En modifiant la fonction coût, on n'observe pas de changement tandis que le fait de modifier la fonction prior contribue à aplanir les variations en réduisant fictivement le nombre de classes de moitié. Une solution pour ne pas avoir à effectuer un nouvel apprentissage à chaque fois que l'on change la fonction coût/prior serait d'effectuer un apprentissage conséquent une bonne fois pour toute, ou alors d'optimiser le nombre de passes et d'utiliser la méthode β (la plus performante).