

## Compte Rendu du TP4

Membre : Hugo 16241007

Arnaud 16241047

Dans ce TP4, on fait les travaux sur le risque de Bayes. De mon point de vue, c'est le cas plus général parce que on prend plus d'attention aux cas plus importants donc il faut y mettre non seulement un poids de probabilité mais aussi un poids de coût quand on estime le risque, autrement dit, l'erreur.

On fait d'abord une comparaison avec le discriminateur avec  $\mu$  et  $\Gamma$  connu. En prenant un prior uniforme et un coût de kroneck inverse, c'est normal d'obtenir les mêmes résultats car on fait les mêmes calculs, mais quand on change le prior et coût, on voit très clairement le risque de discriminateur de Bayes est beaucoup moins que celui de discriminateur avec  $\mu$  et  $\Gamma$  connu. C'est raisonnable puisque le coût de mis-classifier le 2<sup>e</sup> classe est beaucoup plus cher que les autres.

Puis on classifie 10 chiffres comme TP3, mais cette fois-ci on utilise le discriminateur de Bayes et on ajoute une matrice de corrélation. De cette façon on peut savoir le discriminateur confond lesquelles classes. C'est utile quand on analyse les risques en changeant nos prior et coût. Pour la base, on a l'impression que dans le TP3, quand la base de généralisation et la base d'apprentissage suivent des lois différentes, on n'arrive pas à trouver un bon discriminateur car les paramètres qu'il a appris dans la base d'apprentissage n'est plus utile. Mais cette fois-ci, on a bien réussi à trouver un bon résultat. C'est parce que, ce que le discriminateur a appris est  $P(x|\omega_j)$ , ça n'influence pas le discriminateur de Bayes. Donc quand il y a beaucoup des cas à tester, on pourrait gagner du temps.

Quand on passe au cas de réseau de neurones, on voit que au cas où le prior est équiprobable, le discriminateur de Bayes fonctionne moins bien que le discriminateur RN. Et quand on utilise une base d'apprentissage de taille plus grande, on trouve que le tau\_g et R sont tous améliorés. Mais on ne sait pas pourquoi le discriminateur RN $\beta$  fonctionne mieux que le discriminateur RN $\kappa$ , presque pour tous les cas. En fait, de notre point de vue,  $\kappa$  est plus distingué que  $\beta$ , parce que quand  $\eta$  soit positive, on obtient un  $\kappa$  beaucoup plus grand que les cas  $\eta$  soit négatif. Et on a  $\frac{\kappa_i}{\sum_n \kappa_i}$  qui ressemble plus à la fonction softmax donc en théorie il marche très bien pour une multi-classification. Donc ça contredit à ceux que l'on a vu pendant le TP.

Pour proposer une solution qui marche pour tous les cas sans refaire la base d'apprentissage, on pense que d'abord c'est mieux d'avoir une taille autant grande que possible et puis on décide le coût selon le prior de base de généralisation et finalement on propose que on utilise le discriminateur RN $\beta$ . (car il utilise la formule sigmoid)