

1.1 tau_g et R de ses deux méthodes sont égaux et on voit les surfaces discriminants ressemblent. En fait pour alpha choisi le risque de Bayes est égal à la probabilité d'erreur. Comme tau_g est le proba de succès, donc on a 1-R=tau_g. C'est ce qu'on a trouvé.

1.2 Soit matrice de confusion M, $P(w_i | w_j) = M_{ij} / (M_{i1} + \dots + M_{ij} + \dots)$ comme 1.1 on a 1-R=tau_g.

1.3 tau_g pour mu et gamma connus varie un peu (entre ±0.05). C'est une variation normale qu'on a observé d'avant. Mais celui pour Bayes diminue et on voit que la surface pour C2 manger un peu celle de C1. Comme on augmente a12=100, il va prendre plus d'importance d'éviter C2 de mal placer C1. Et comme on a ajouté cette importance, c'est normal de perdre un peu de tau_g parce que si l'on se concentre sur une importance partielle c'est normal de perdre la totalité.

1.4 On a R diminue pour Bayes donc tau_g augmente.

explication 1 : pour Bayes, un cas équiprobable est plus difficile à distinguer qu'un cas où proba d'une classe est dominante.

explication 2 : voir définition de $R = \sum_{i=1}^C \sum_{j=1}^C \alpha_{ij} P(\omega_j) P(\omega_i | \omega_j)$ | pour cas équiprobable, on a P(wj) sont presque tous égaux à 1/3(moyenne) et P(w_i | w_j) comme on a trouvé par matrice de confusion, ils sont presque tous égaux et à valeur moyenne. Mais pour un cas dominant, on a 2 P(wj) est petit (souvent plus petit que P(wj) P(w_i | w_j) pour cas équiprobable). Pour le reste, comme on voit P(w_i | w_j) est souvent plus petit que P(wj) P(w_i | w_j) pour cas équiprobable. Donc on trouve pour le total R devient plus petit.

2.1 les zones presque noir dans l'image des coefficients de corrélation entre les images de chiffres quand i≠j, qui correspond aux nombres >40 pour i≠j, c'est pour les mauvaises estimations.

2.2 On sait que pour Bayes simple qu'on sait mu et gamma, le risque sont tous égaux pour tous Papp. Mais pour linéaire et quadratique il faut prendre plus d'apprentissage pour atteindre stable. De plus pour quadratique, comme il y a 42 dimensions, il faut prendre beaucoup d'apprentissage que celui de linéaire.

2.3 1 : c'est possible parce que pour R P(w_i | w_j) c'est ce qu'on peut apprendre une fois puis on n'a pas besoin de recommencer. Et les autres deux ne dépendent pas d'apprentissage.

2. toujours comme 1 : ajouter une importance pour aij on perd tau_g et et créer un classe dominant on gagne tau_g

3.1 RN Bayes est un droite avec R plus petite. RN_beta est moins concentré avec R plus grande et plus de valeur différent. RN_gaba est pas très concentré avec R au moyenne plus grande que les deux autres et possède beaucoup plus de valeur différent. **Je ne suis pas sur si c'est parce que c'est à cause de le choix de gaba et beta : la difference entre les beta_i est plus petite que la difference entre les gaba_i. Et cette differents estimation de proba cause la difference mathematiquement qui influence finalement de R.(besoin d'explication)**

3.2 c'est toujours comme 1 et 2.3.2

3.3 **pas sur : gaba et beta sont definis par la sortie, et on peut le garder une fois puis on l'applique pour tous les cas.(besoin d'explication)**