# II. Réduction de la dimension

Pour réduire la dimension de notre espace () nous pouvons utiliser deux méthodes :

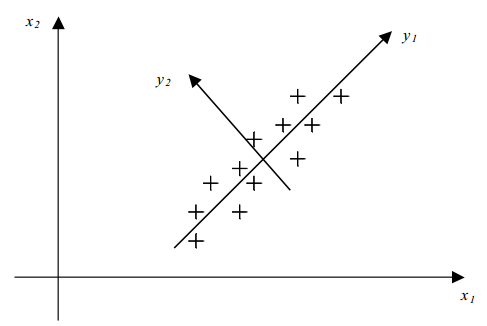
* Projection
* Sélection

La projection c’est garder les caractéristiques les plus précises. Quelles sont les dimensions qui donnent le plus d’information sur notre objet ? La sélection c’est plutôt choisir celles qui discriminent le mieux les caractéristiques. Quelles sont les meilleurs composantes parmi les qui font le plus diverger les classes.

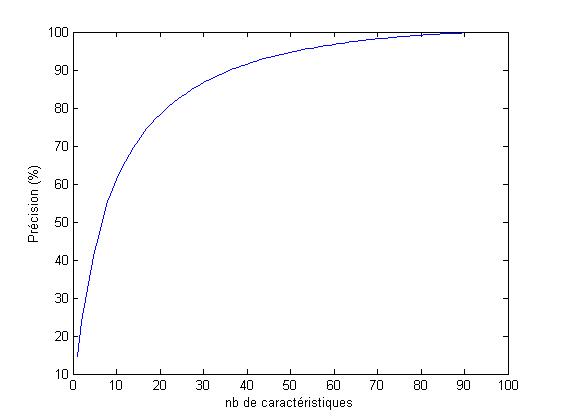
## Projection des caractéristiques

Il existe beaucoup de méthodes de projection des informations d’un objet. Nous nous concentrerons sur une méthode qui est l’analyse par composante principale (ACP).   
L’analyse en composantes principales est une méthode statistique qui consiste à extraire les vecteurs propres et les valeurs propres de la matrice de covariance, calculée à partir de l’ensemble de données d’apprentissage. L’intérêt de cette technique est de rechercher les axes principaux de l’ensemble des données qui ont la plus grande variation (valeur propres). En ayant cette information de vecteur propre on peut ainsi projeter nos données selon les axes fournissant le plus d’informations et composé de caractéristiques décorrélées.

Par exemple pour une classe donnée, on peut retrouver 2 vecteurs propres et on choisit l’axe qui possède le plus d’informations (la plus grande variance des données).



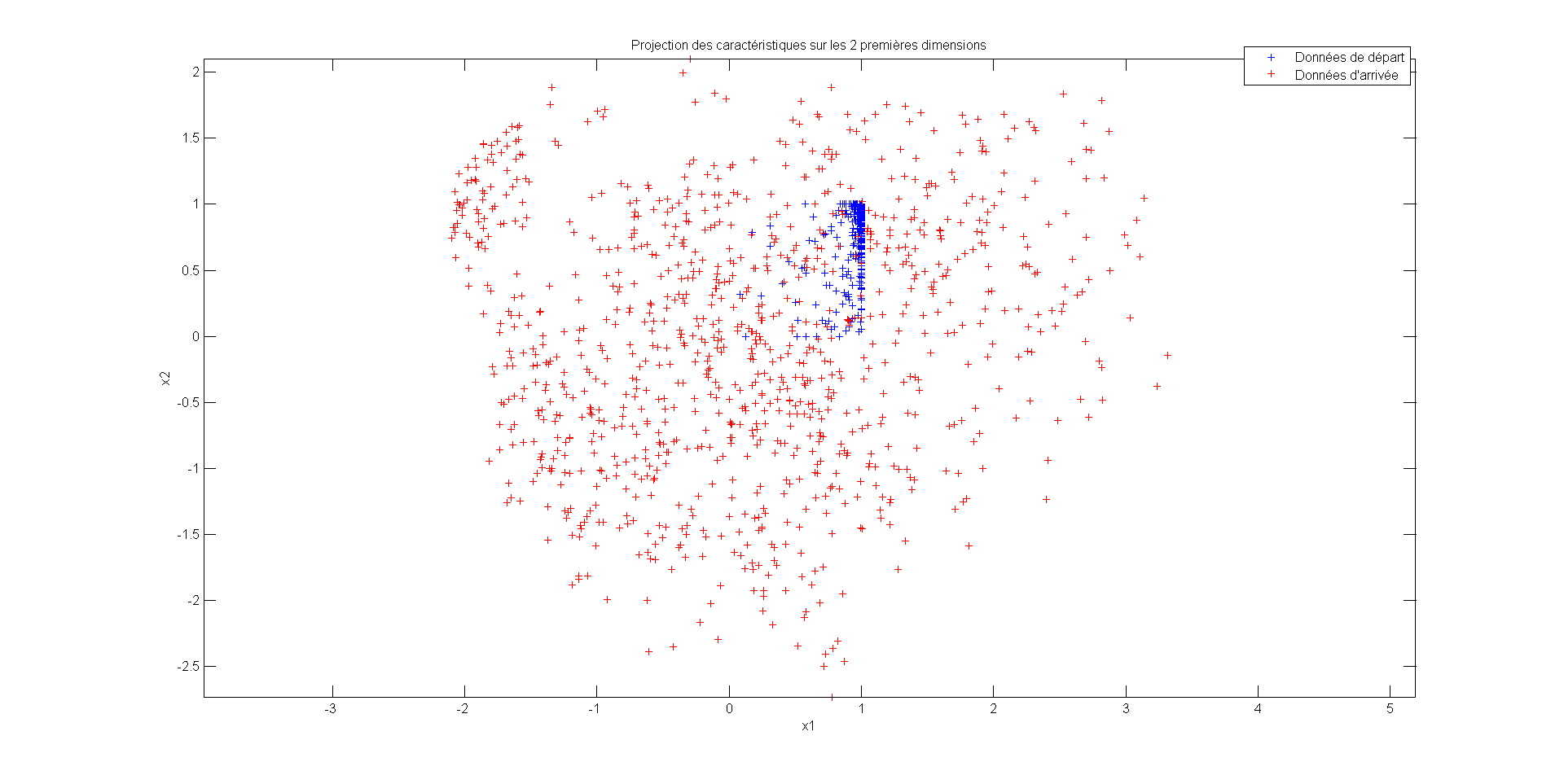
Nous avons donc calculé les différentes valeurs et vecteurs propres pour notre base de données totale. Voici un graphique représentant le pourcentage d’informations en fonction du nombre de caractéristiques gardées :



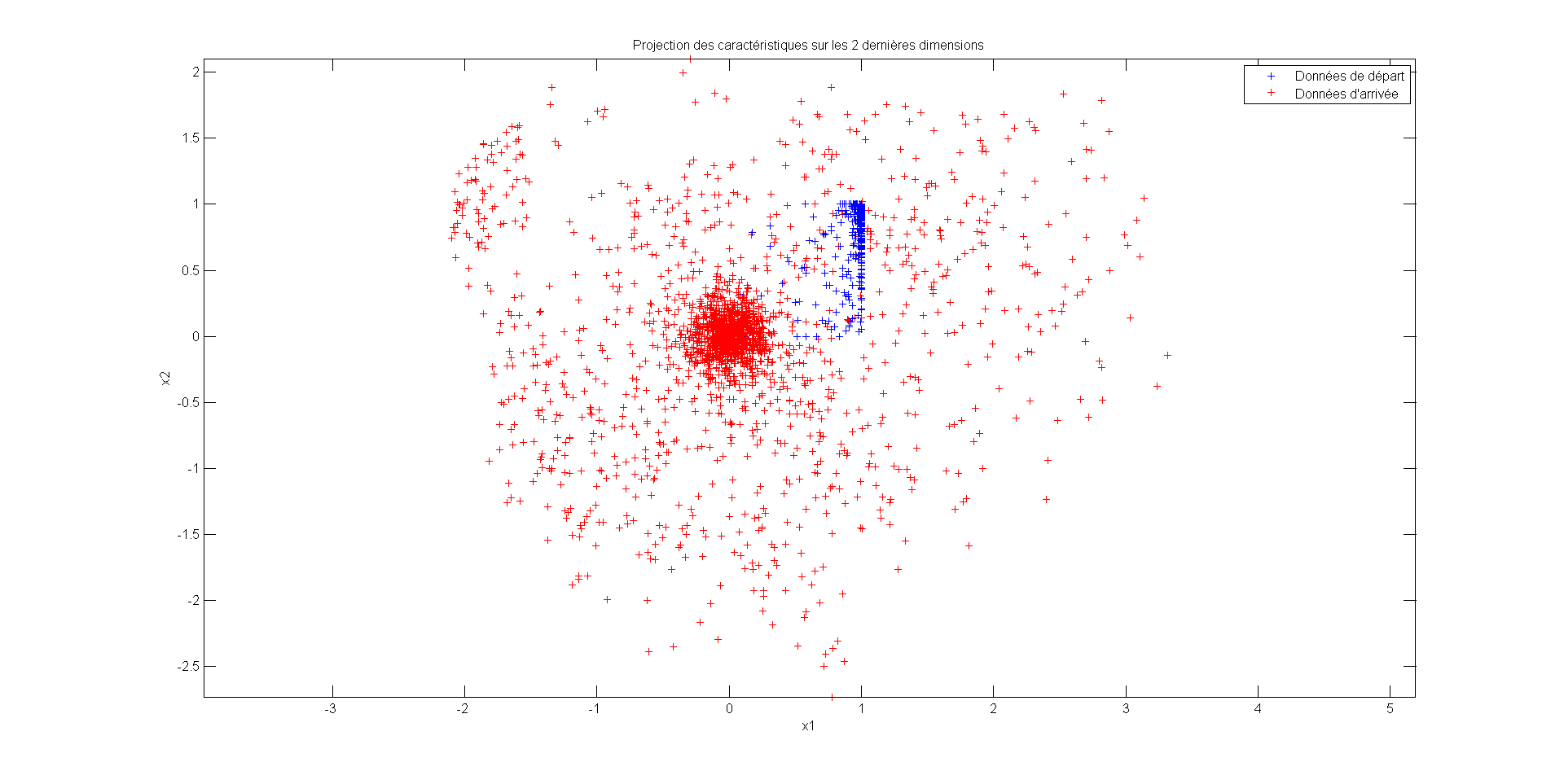
Nous nous fixons un objectif qui est d’obtenir :

C’est-à-dire que la somme de nos valeurs propres choisies doit correspondre à un poids de parmi toutes les valeurs propres. 95% est beaucoup utilisé en statistiques et est une règle générale pour le choix du taux de rejet à . En appliquant cette règle, nous avons décidé de garder les 52 meilleurs axes parmi les 100 disponibles. Il est nécessaire maintenant de projeter les données de notre **base de test** dans notre nouvel espace défini par les vecteurs propres précédemment choisis. Sans oublier bien évidemment de normaliser.

Nous avons donc qui possède 52 caractéristiques. En guise de résultat, voici un exemple en prenant les 2 premières dimensions et en projetant les données dans les 2 meilleurs vecteurs propres. On voit que l’on a une répartition beaucoup plus variable que les données de départ en bleu.



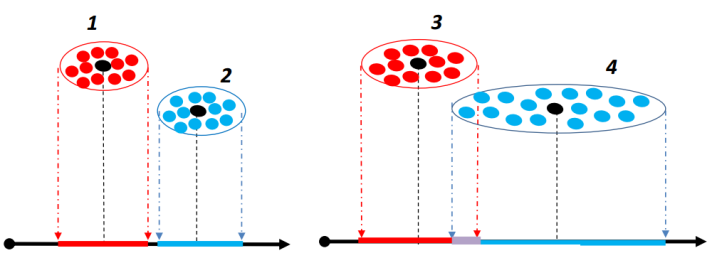
En projetant sur les 2 derniers vecteurs propres, on voit au contraire que la variabilité est moins élevée et les données sont beaucoup plus concentrées. On remarque donc que les 2 premiers vecteurs propres sont plus intéressants que les 51 et 52èmes (dans le cas où l’on en a choisi 52).



Par contre il faut faire attention à une chose, c’est qu’il ne faut pas généraliser le cas. Autant pour les 2 premières dimensions les résultats sont concluants, mais ça ne veut pas dire que l’on aura de bons résultats pour les autres. Il ne faut pas oublier que l’ACP fait perdre de l’information (dans notre cas 95% de l’information).

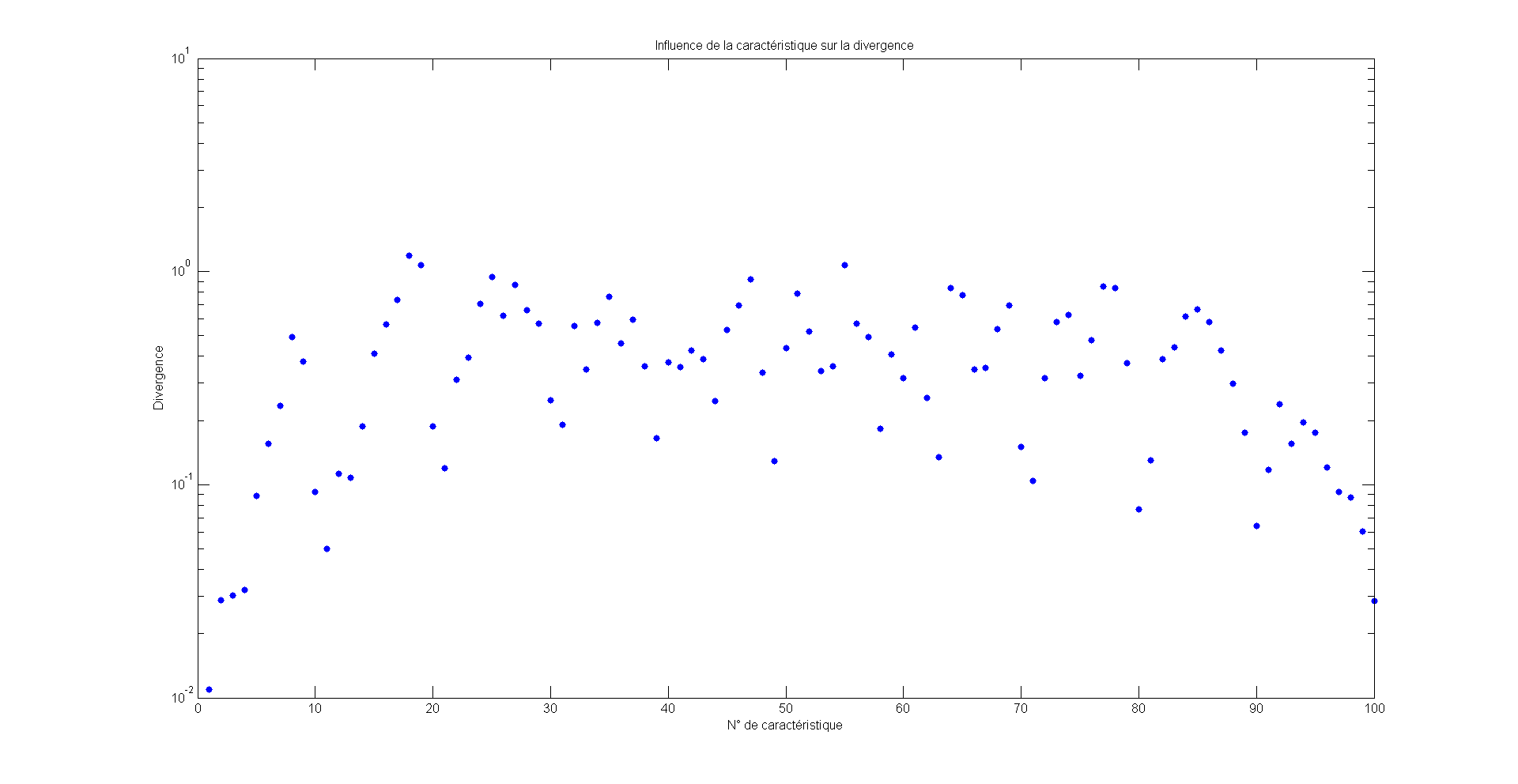
## Sélection des caractéristiques

Une fois que l’on a notre information sur les axes contenant le plus d’informations, on peut sélectionner les meilleurs axes. Pour cela, on va calculer les divergences entre les classes, ainsi si la divergence est bonne il s’agit d’une bonne caractéristique.

  
1 et 2 divergent bien contrairement à 3 et 4.

Pour calculer la divergence entre les classes, nous utiliserons le **critère de fisher** qui a déjà été utilisé au préalable.

Pour sélectionner les dimensions, il est nécessaire de travailler sur chacune des dimensions. Ainsi sur *MATLAB* on a une matrice 3D avec 100 tableaux de divergence (un tableau pour chaque caractéristique). Pour obtenir l’information de la divergence entre les classes et ce pour chacune des caractéristiques, on effectue la moyenne de la matrice de divergence. On obtient ainsi les valeurs suivantes pour chaque dimension :

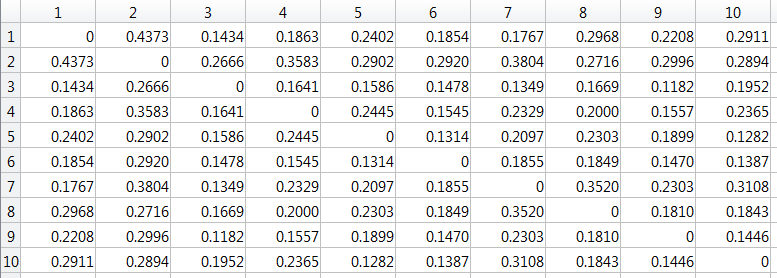


On considère que l’on a une bonne divergence à partir de (au sens de Fisher). En effet à partir de , on ne retrouve plus de recouvrement avec en plus une petite marge entre les 2 classes (pour être sûr). L’application de ce critère nous permet de sélectionner 33 caractéristiques parmi les 100 ce qui est raisonnable. Dans l’image de gauche ci-dessous, les régions de rétine coloriées en noir correspondent donc à celles qui sont les plus discriminantes après calcul. L’image de droite quant à elle donne le niveau d’importance de chaque région de l’image, plus le pixel est sombre, plus la région est importante.   
On a donc une sélection plutôt logique au vu des images de la base de test. En effet on remarque entre autre que les côtés ne sont vraiment pas sélectifs car on a centré l’image et donc les pixels sont plutôt concentrés au milieu.

# Conclusion

Alors que la sélection va permettre de mieux discriminer les données rentrantes, la projection permet, elle, d’obtenir plus d’informations en utilisant un nouvel espace pour nos données. Dans le premier cas, on a sélectionné 33 caractéristiques parmi les 100, dans la projection on en a sélectionné 52.   
Avec la sélection, on travaille sur des zones connues puisque l’on ne change pas la base dans laquelle nous travaillons. La projection nous projette dans une nouvelle base qui n’a plus vraiment de sens physique (les caractéristiques ne sont plus les régions de la rétine). Par contre l’ACP permet d’obtenir une base contenant plus d’information que celle de départ. Malheureusement avec une mauvaise divergence entre les classes comme le montre le tableau suivant (pour la meilleure caractéristique, plus grande valeur propre) :



Nous sommes donc dans le cas où les 2 méthodes se valent, il pourrait alors être intéressant de combiner les 2 techniques pour avoir les avantages de l’un et l’autre. Un dernier point à prendre en compte au niveau de la sélection et la projection des caractéristiques est la malédiction de la dimensionnalité (« curse dimensionnality »). Elle met en garde le concepteur du système à utiliser un nombre limité de caractéristiques quand seulement un faible nombre d’échantillons d’apprentissage est disponible.   
En général on dit que la taille de notre échantillon doit être proportionnelle à la dimension.

L’idéal étant de minimiser au maximum notre dimension est de maximiser la taille de l’échantillon. Dans notre cas, il faut donc une dimension ce qui est le cas pour les deux techniques que nous avons utilisées. Attention cependant au fait que le taux de recouvrement augmente avec le nombre de caractéristiques et il y a un compromis.