

Machine learning  
ET4 informatique  
Projet de reconnaissance d'images

CLAUDE Philippe, DIONGUE Abdoulaye, JACQUEMIN Félix

March 26, 2019

## Introduction

Ce document représente le rapport du projet, reconnaissance d'images, réalisé pendant le semestre 8 dans le cadre du cours "Option découverte, Machine learning". Le groupe réalisant le projet est composé de CLAUDE Philippe, DIONGUE Abdoulaye, JACQUEMIN Félix.

L'objectif du projet est de déterminer les images mal classées parmi plusieurs en fonction d'un ensemble de test. Pour ceci, on utilise plusieurs méthodes afin de déterminer celle qui a la meilleure performance en temps et taux d'erreur.

Dans ce rapport, on présentera les choix de classifieurs que nous avons utilisés. Ensuite, effectuer une comparaison entre ces différents classifieurs à travers des figures illustrant le résultat de chaque méthode.

## Choix de classifieurs

Tous les classifieurs présentés ci-dessous seront appliqués sur les données avant et après l'application du ACP. Ce dernier consiste à réduire la taille des données tout en gardant toutes les informations nécessaires pour appliquer un classifieur sur ces données.

### Classifieur à distance minimum

Ici, on regroupe les images ayant la même classe en calculant le barycentre de ces images. Ensuite, on calcule pour chaque image sa classe qui est la distance minimale par rapport à sa classe donnée. Ainsi, on peut déterminer si l'image est mal classée en comparant sa classe calculée avec la classe donnée.

### Support Vector Machines (SVM)

Les SVMs sont des classifieurs supervisés. Ils permettent de déterminer la classe d'une image sur un ensemble d'images. Pour ce faire, ils appliquent une fonction sur les images en entrées afin d'obtenir une fonction en sortie qui doit correspondre à la bonne classe de l'image.

### Les avantages des SVMs

- Efficace dans les espaces de grandes dimensions
- Toujours efficace dans les cas où le nombre de dimensions est supérieur au nombre de données à classifier.
- Utilise un sous-ensemble de points d'apprentissage dans la fonction de décision, ce qui en fait une mémoire efficace

## Les SVMs utilisés

**Linear SVC :** Il est semblable au SVC avec kernel linear mais il est implémentée avec linéarité plutôt que libsvm. Cette SVM prend en charge les données denses et fragmentées.

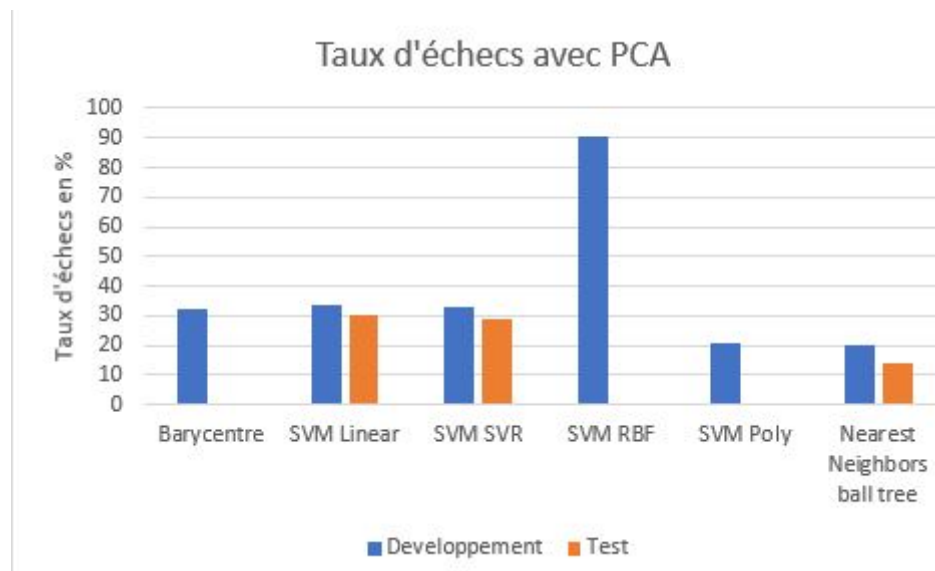
**SVC avec kernel linear, rbf et poly:** Ici, les SVMs sont implémentées avec libsvm qui est une bibliothèque du SVM.

## Les plus proches voisins : neighbors

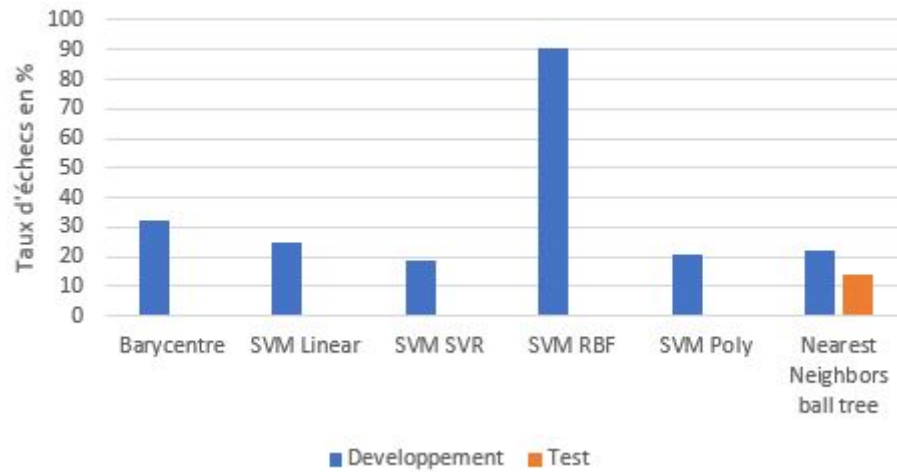
Le principe de ce classifieur est de considérer un certain nombre de voisin pour une image et renvoyer la distance entre l'image et chacune de ses voisins. Ensuite, un modèle est ajusté en utilisant l'image en entrée et sa classe cible. Ainsi, il pourra prédire la classe de l'image afin de savoir si elle est mal classée ou non.

## Comparaison des classifieurs

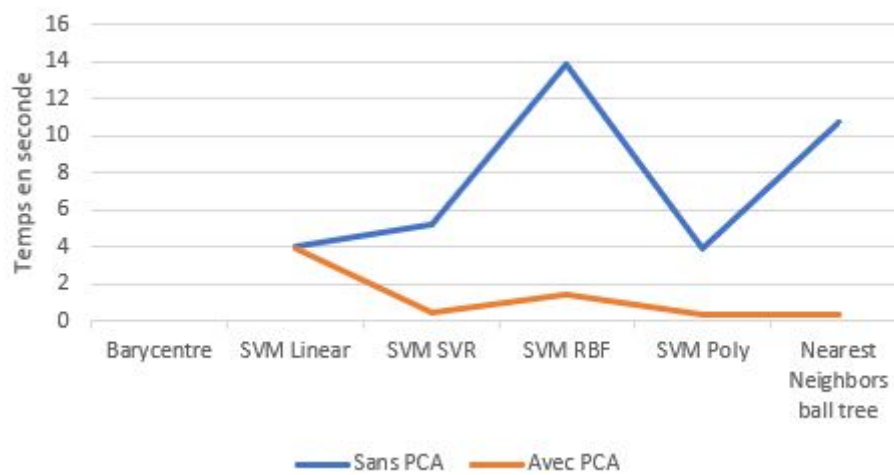
Dans cette deuxième partie nous allons voir les différents classifieurs, le barycentre est un classifieur moyen avec une performance à 33 % ce qui est déjà très bien pour un algorithme simple à implémenter. Pour le moment nous allons regarder ces algorithmes seulement sans la ACP.



### Taux d'échecs sans PCA

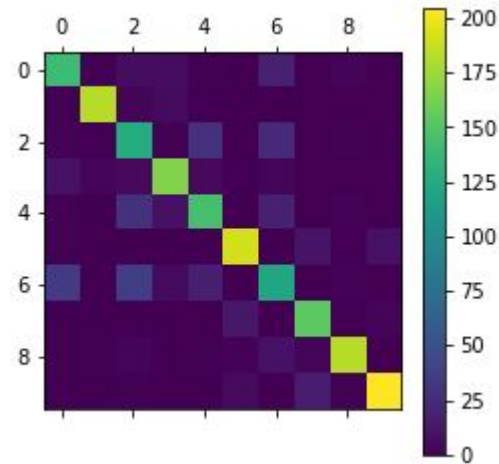


### Temps d'exécution

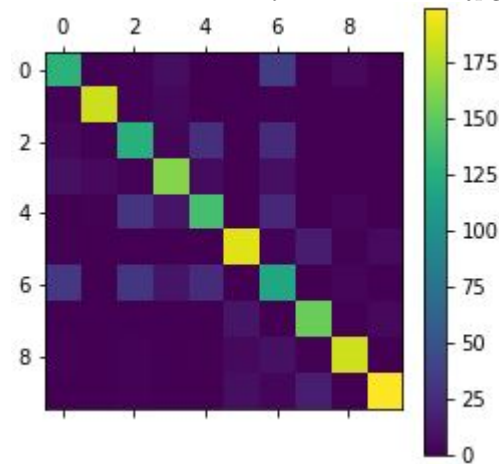


Celui qui obtient le meilleur score est la SVM SVR avec un noyau linéaire car le choix du point central est mieux réalisé que sur la SVM Linéaire.

de confusion SVM SVR sans PCA dev.jpg

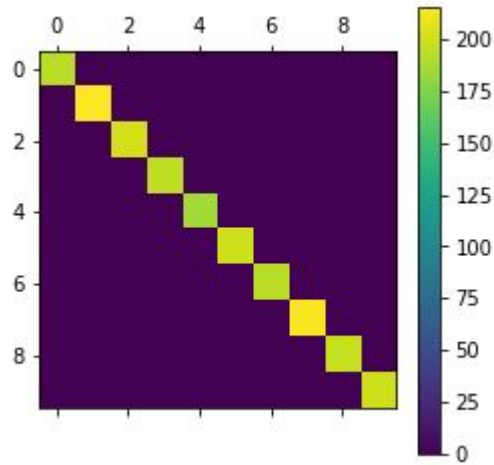


de confusion SVM Poly avec PCA dev.jpg

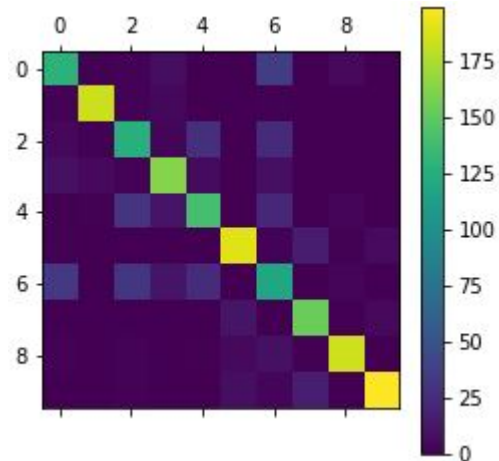


Cependant avec une ACP celui qui obtient le plus petit taux d'échecs est la SVM Polynomiale car elle intègre mieux la perte des dimensions avec des courbes de décisions plus elliptique et non des droites qui accepte moins le changement vers des données trop différentes. D'ailleurs les matrices de confusion avec la ACP nous confirment ce choix pour la SVM Polynomiale qui démontre que nous avons bien plus d'éléments correspondant sur la diagonale.

de confusion SVM RBF avec PCA test.jpg



de confusion SVM Poly avec PCA dev.jpg



Un autre algorithme se défend bien c'est celui des plus proches voisins qui à un taux d'échecs sur les données d'entraînement quasiment aussi important que pour les données de développement mais il reste bon pour la prédiction plutôt stable même si l'on réduit les dimensions des données d'entrées.

Cependant avec nos deux facteurs le taux d'échecs ainsi que le temps d'exécution la SVM Polynomiale gagne haut la main avec un temps proche de 4 secondes sans ACP et quasiment nul avec.

Nous n'avons pas utilisé l'ensemble de test car nous ne pouvions pas calculer un taux d'erreurs ce qui nous est indispensable pour savoir si un classifieur est viable ou non, en somme sans vérification possible, nous ne pouvons pas juger si ce classifieur est correct ou non.