

# Reconnaissance de visages

Marie Szafranski

26 mars 2015

## Préambule

**Énoncé en ligne.** L'énoncé contient un certain nombre de références en ligne. Pour vous faciliter l'accès à ces ressources, vous trouverez cet énoncé à l'adresse [http://www.ensiie.fr/~szafranski/downloads/MOST/14\\_15/projet/projet\\_ias.pdf](http://www.ensiie.fr/~szafranski/downloads/MOST/14_15/projet/projet_ias.pdf).

**Rendu.** Ce projet est à réaliser en binôme. Vous devrez déposer une archive au format zip ou tar contenant :

- Un compte-rendu de *13 pages maximum*<sup>1</sup> au format pdf,
- Un répertoire contenant vos codes accompagné d'un fichier txt ou pdf décrivant la façon dont s'organise votre code.

Cette archive devra être remise au plus tard le 24 avril 2015 sur le dépôt most-ias-projet-2015 du serveur <http://exam.ensiie.fr>. Vous ne devez déposer qu'un seul compte-rendu par binôme. Le non-respect d'une des consignes entraînera des points de pénalité. Tout travail qui ne sera pas rendu au bon format (zip ou tar pour l'archive, pdf pour le compte rendu et txt ou pdf pour la description de l'organisation du code) sera considéré comme manquant. Un projet qui ne sera pas rendu dans les délais se verra attribuer la note de 0.

## 1 Projet

**Données.** Les données sont disponibles sur [http://www.ensiie.fr/~szafranski/downloads/MOST/14\\_15/projet](http://www.ensiie.fr/~szafranski/downloads/MOST/14_15/projet). La base de données contient 165 images (en niveaux de gris) de 15 personnes différentes. Pour chaque personne, on a 11 images : une par expression. Vous pouvez trouver plus d'information sur ces données sur le site <http://vision.ucsd.edu/content/yale-face-database>.

**Objectif.** Vous devez mettre en place une procédure d'apprentissage pour tester et comparer trois méthodes de classification sur les données présentées ci-dessus :

- Les *k plus proches voisins*, [À faire par les deux membres du binôme]
- Les SVM, [À faire par un des deux membres du binôme]
- Le *boosting*. [À faire par l'autre membre du binôme]

**Codes.** Vous pouvez utiliser certaines implantations des méthodes fournies dans le cadre des TP précédents ou celles de Matlab directement [4, 2, 3]<sup>2</sup>. Vous pouvez également utiliser d'autres langages si vous le souhaitez. Par exemple, la boîte à outils Scikit-learn sous python est fiable et très complète [1]. Le logiciel scientifique R que vous utilisez dans le module de *régression avancée* est également fiable et (trop<sup>3</sup>...) complet [5, 6]. Par contre, si vous utilisez d'autres codes trouvés sur Internet, vous devrez vous assurer qu'ils sont fiables.

1. La page de présentation, la table des matières et les éventuelles annexes doivent être intégrées dans ce compte. Des points de pénalités seront attribués pour chaque page supplémentaire.

2. Les références concernent la version 2015 de Matlab, mais vous pouvez aussi consulter l'aide directement depuis Matlab.

3. Il y a parfois plusieurs implantations d'une même méthode.

## 2 Rapport

Le rapport devra contenir :

- Une introduction [~ 1 page]  
*L'introduction devra décrire les objectifs abordés, présenter les données utilisées et donner un plan détaillé du rapport. De plus, vous devrez préciser la répartition des tâches dans le binôme (qui fait les SVM et qui fait le boosting).*
- Une description de la stratégie utilisée pour comparer ces méthodes [~ 1 page]  
*Il s'agit de décrire la procédure d'apprentissage ainsi que la procédure de calcul de l'erreur sur l'ensemble des classes.*
- Une présentation claire et synthétique des résultats obtenus par méthode [~ 1 à 2 pages par méthode]  
*Pour chaque méthode, vous donnerez l'implantation utilisée ainsi que la grille des paramètres explorés.*
- Une comparaison des résultats sur l'ensemble des méthodes [~ 1 page]
- Une conclusion [~ 1 page]
- Une bibliographie [~ 1 page]  
*Si vous vous inspirez de supports, même disponibles en ligne, vous **devez** citer vos références.*

**Forme du rapport.** Le nombre de page du rapport pour chaque partie est donné à titre indicatif. Néanmoins, le rapport ne doit *en aucun cas* excéder 13 pages. Apprenez aussi à utiliser correctement votre traitement de texte : table des matières réalisée en fonction de la gestion des titres ; cohérence dans les styles des titres, des puces, des textes, etc. tout au long du rapport ; utilisation d'un correcteur orthographique ; etc.

**Présentation des résultats.** Les résultats doivent être présentés sous forme de tableaux ou de graphiques synthétiques. Chaque tableau et chaque graphique doit être numéroté et référencé dans le texte principal. Sous chaque tableau et chaque graphique, il doit y avoir une légende accompagnée d'un texte expliquant quels sont les résultats présentés. De plus, sur chaque graphique, il doit y avoir un titre, les axes et les courbes doivent être identifiés, les courbes doivent être lisibles si elles sont imprimées en noir et blanc. Les codes devront être en annexe. Le non-respect d'une de ces consignes entraînera des points de pénalités.

**Code dans le rapport.** Si vous souhaitez intégrer des extraits de (pseudo-) code pertinents (par exemple la description de la procédure d'apprentissage) , ils doivent se trouver en annexe et doivent être commentés.

## 3 Quelques remarques

1. Les notes ne seront pas données en fonction des taux d'erreurs trouvés sur l'une des méthode (il est possible qu'une méthode fonctionne moins bien que les autres) mais en fonction de la façon dont :
  - La procédure d'apprentissage sera mise en œuvre sur l'ensemble des méthodes,
  - La qualité de la présentation des résultats,
  - La pertinence des commentaires.
2. Gardez à l'esprit que, quelle que soit l'implantation utilisée, vous devez faire en sorte de comparer des choses cohérentes à travers les méthodes. En particulier, vous devez prendre ça en compte en définissant votre procédure d'apprentissage.

## Références

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, and (...). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011. URL <http://scikit-learn.org/stable/>.
- [2] Matlab. SVM, R2015a. URL <http://www.mathworks.fr/fr/help/stats/support-vector-machines.html>.
- [3] Matlab. Boosting, R2015a. URL <http://www.mathworks.fr/fr/help/stats/boosting.html>.
- [4] Matlab. knn, R2015a. URL <http://www.mathworks.fr/fr/help/stats/nearest-neighbors.html>.
- [5] R. Machine learning for R, 2007. URL <http://cran.r-project.org/web/views/MachineLearning.html>.
- [6] R. rminer, 2012. URL <http://cran.r-project.org/web/packages/rminer/index.html>.