

Contrôle continu : Fondements du Machine Learning

L3 IM2D

Janvier-Février 2025



Instructions générales

- Version courante : 17 janvier 2025.
- La version la plus récente du projet (mise à jour en cas de changement) est disponible à l'adresse <https://www.lamsade.dauphine.fr/~croyer/ensdocs/FML/CCFML.pdf>.
Historique des changements
 - 2025.01.17: Ajout de référence pour effectuer la SVD.
- Le travail décrit ci-dessous peut être réalisé en binôme ou individuellement.
- Les réponses aux différentes questions ainsi que les sources Python devront être envoyées au format ZIP à clement.royer@lamsade.dauphine.fr. Le nom de l'archive devra comporter le(s) nom(s) de chaque étudiant(e) impliqué(e) dans le projet.
- La date limite de rendu est fixée au **7 février 2025 AOE** (Anywhere On Earth).

Introduction

Ce court projet se base sur les deux TD/TP réalisés en séance lors du cours, dont les corrections sont disponibles en ligne. Il s'agit de reprendre ceux-ci en apportant quelques modifications, et d'analyser brièvement les résultats à travers les questions ci-dessous.

Retour sur le TP 1 : ACP

Question 1 Choisir deux des images parmi celles extraites au début du TP, et comparer leur reconstruction par SVD tronquée avec celle par composantes principales.

Il est recommandé d'utiliser une implémentation existante de la SVD, par exemple la fonction `svdvals` de la bibliothèque `scipy`.

- Quelle technique fournit la meilleure reconstruction en termes de norme de Frobenius ?
- Commenter sur la qualité visuelle des reconstructions de chaque image.

Question 2 Le TP considérait une base d'images fixée parmi la bibliothèque considérée. Reprendre le TP en sélectionnant 9 images d'un même individu comme base de travail.

- La reconstruction des individus est-elle meilleure avec cet ensemble ?
- Que devient le résultat de la partie 3 lorsque l'on introduit une nouvelle image du même individu ou d'un autre individu ?

Retour sur le TP 2 : Régression linéaire

Question 3 Reprendre la partie 1 du TP en changeant les dimensions n et m en $n = 40$ et $m = 10$.

- Quelle erreur obtient-on avec et sans a priori ?
- En augmentant le paramètre σ définissant les données, comment évolue l'erreur de la question précédente ?

Question 4 Reprendre la partie 2 du TP en ajoutant un a priori gaussien sur le modèle de la forme $\mathcal{N}(\mathbf{0}, \frac{1}{\lambda} \mathbf{I}_n)$. Comparer les résultats avec et sans a priori pour diverses valeurs de λ .

- Quel est l'impact de la régularisation sur l'erreur d'apprentissage pour une dimension fixée ?
- Comment évolue la meilleure valeur de d en fonction de λ ?