

GSI-ACAD (2018-2019)
 Machine Learning
 TP 1 : modèles linéaires pour la régression

Remarques préliminaires :

- Ce TP est noté.
 - Vos pouvez travailler en monôme ou en binôme.
 - Votre travail est à rendre, au plus tard, une semaine après la séance de TP. Par exemple, le travail relatif au TP du vendredi 21/09 8h devra être rendu avant le vendredi 28/09 8h.
 - Ce qu'il faut rendre : tous vos codes sources + compte rendu, par mail :
moncef.hidane@insa-cvl.fr.
1. Pour vous refamiliariser avec Matlab, prenez le temps (1h) de parcourir les documents qui se trouvent ici : <https://dspace.mit.edu/bitstream/handle/1721.1/58528/6-094-january-iap-2009/contents/lecture-notes/index.htm>.
 2. **Régression polynomial 1D.** On reprend l'exemple de la fonction $x \mapsto \sin(2\pi x)$ dans l'intervalle $[0, 1]$.
 - (a) Écrire la fonction `[x,t] = generate_training_set(N, std)`. Cette fonction génère un ensemble d'apprentissage $\{(x_1, t_1), \dots, (x_N, t_N)\}$ où les x_i sont générés selon une loi uniforme dans $[0, 1]$ et $t_i = \sin(2\pi x_i) + \epsilon_i$ ou ϵ_i est une réalisation d'une gaussienne centrée d'écart-type `std`.
 - (b) Écrire la fonction `w = pol_ls_fit(x, t, M, opt)`. Cette fonction ajustera le paramètre **w** d'un polynôme de degré M en minimisant la somme des carrés des erreurs sur la base d'apprentissage. Le paramètre `opt` sera une structure qui permettra de déterminer la méthode de résolution (pseudo inverse ou descente de gradient, nombre d'itérations, critère d'arrêt et pas de descente si descente de gradient).
 - (c) Validez la fonction `pol_ls_fit` en comparant la sortie à celle produite par la fonction `polyfit` de Matlab.
 - (d) Générez la figure de la diapositive 18.
 - (e) Générez le tableau de la diapositive 20.
 - (f) Écrire la fonction `w = pol_ridge_fit(x, t, M, lambda, opt)`. Cette fonction met en oeuvre une régression ridge par descente de gradient. `opt` est une structure qui contient les paramètres de l'algorithme de descente de gradient.
 - (g) Générez la figure à droite de la diapositive 21.
 - (h) Générez la figure à droite de la diapositive 30, illustrant la décomposition biais-variance.
 - (i) Générez une figure similaire à celle de la diapositive 43. Au lieu d'échantillonner **w** selon la loi a posteriori, on échantillonnera directement t selon la distribution prédictive.
 3. **Prédiction de la consommation d'une voiture.**
 - (a) Chargez la base `carsmall` dans Matlab. En essayant un modèle linéaire de régression (vous avez le choix du type et nombre de fonctions de base), essayez de prédire la consommation MPG (miles per gallon) à partir du vecteur (`Weight Horsepower Cylinders Model_Year`). La validation se fera en calculant la moyenne sur 10 expériences de la moyenne quadratique (E_{RMS}) sur 20% de la base, tirés aléatoirement (les 80% restant servant à l'apprentissage).