

# **Mini-projet : Régression linéaire (Prédiction de la flexion des doigts à partir des signaux électrocorticographiques)**

Le but de ce mini-projet est de découvrir l'apprentissage de fonction de prédiction linéaire et non linéaire par la méthode des moindres carrés et des moindres carrés régularisés. L'évaluation de cette prédiction est opérée sur des bases de données réelles (Biomédical, Energie, Finance, etc.)

## **Description des données**

Pour ce mini-projet, on considère le jeu de données 4 de la compétition BCI IV dont le but est de prédire la flexion des doigts du sujet à partir de mesures des signaux électrocorticographie (ECOG).

Les trois sujets de la base de données sont des patients épileptiques ayant des capteurs ECOG en place pour des raisons médicales. Les sujets sont équipés de gants permettant de mesurer la flexion de chacun de leurs doigts. Il est demandé aux sujets de bouger un de ses doigts avec un ordre visuel.

On sélectionne pour ce min projet les signaux d'apprentissage du sujet 3 et on essayera de prédire la flexion de son pouce au cours du temps. Le signal mesuré consiste en 10 minutes d'enregistrement à 1000Hz sur 64 électrodes. Il a été filtré avec un filtre passe bas afin d'éliminer les bruits d'instrumentation et les artefacts de la surface électromyographique. Ensuite le signal est sous-échantillonné de manière à obtenir tenir un signal temporel échantillonné à la cadence de  $F_e = 50\text{Hz}$ . Le signal cible (mouvement des doigts) est également sous-échantillonné et uniquement le canal contenant le mouvement du doigt est conservé. Finalement une détection grossière de mouvement est effectuée et seuls les instants temporels correspondant à un mouvement sont conservés.

Le fichier "sub3\_comp.mat" contient les variables suivantes :

Xall Matrice contenant n exemples de d variables.

Yall Vecteur contenant les valeurs à prédire.

$F_e$  fréquence d'échantillonnage des deux signaux.

## Travail demandé :

### 1) Visualisation des données

- a. Charger les données en mémoire, découvrir et visualiser les signaux de la base de données.
- b. Découper les données en un ensemble d'apprentissage et un ensemble de test. Attention ce sont des données temporelles, il ne faut donc pas intervertir l'index des exemples.

### 2) Régression des moindres carrés

- a. Poser le modèle et estimer les paramètres à l'aide d'un programme basé sur la minimisation de la fonction cout et la descente du gradient.
- b. Calculer et afficher sur un graphique la fonction cout à chaque itération.
- c. Tester différentes valeurs de pas dans la descente du gradient. Conclure
- d. Vérifier le résultat analytiquement.
- e. Prédire le mouvement des doigts du sujet pour les signaux ECOG d'apprentissage et de test. Mesurer la performance dans les deux cas. Que remarquez vous ?

### 3) Régression Ridge

- a. Estimer les paramètres de la fonction linéaire dans le cadre de la régression ridge.
- b. Prédire les signaux cibles sur les exemples d'apprentissage et de test, pour différentes valeurs du paramètre de régularisation  $\lambda$ . Sélectionner le  $\lambda$  permettant d'obtenir les meilleures performances en prédiction.

### 4) Optionnel

Interprétation et sélection de variables : Une fonction linéaire a l'avantage d'être interprétable. En effet le vecteur  $\theta$  contient des coefficients correspondant à chaque variable dans X. L'amplitude de ces coefficients donnera donc l'impact de chaque variable dans la fonction de décision.

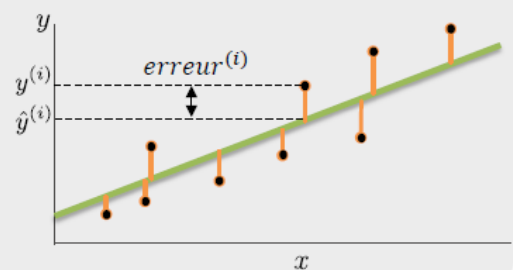
- a. Visualiser la valeur absolue des coefficients du vecteur  $\theta$ . Repérer les capteurs les plus importants dans la fonction de prédiction.

- b. Sélectionner uniquement les variables les plus importantes et refaire l'apprentissage. Évaluer les performances en utilisant uniquement ce sous-ensemble de variables.
- c. Visualiser en 3D les exemples d'apprentissages avec l'hyperplan définissant la fonction (sur les deux variables les plus importantes). Refaire la même chose avec les exemples de test.

## 5) Evaluation des performances en prédiction

Metriques	Equations
Erreur Moyenne (MSE)	$MSE = \frac{1}{n} \sum_{i=0}^n (y^{(i)} - \hat{y}^{(i)})^2$
Erreur quadratique moyenne (RMSE)	$RMS = \sqrt{\frac{1}{n} \sum_{i=0}^n (y^{(i)} - \hat{y}^{(i)})^2}$
Erreur absolue moyenne (MAE)	$MAE = \frac{1}{n} \sum_{i=0}^n  y^{(i)} - \hat{y}^{(i)} $
Coefficient de determination ( $R^2$ )	$R^2 = 1 - \frac{\sum_{i=0}^n (y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=0}^n (y^{(i)} - \bar{y})^2}$

$y^{(i)}$  : les valeurs des données  
 $\hat{y}^{(i)}$  : les valeurs prédites  
 $\bar{y}$  : la Moyenne des données,  $\bar{y} = \frac{1}{n} \sum_{i=0}^n y^{(i)}$   
 $n$  : dimension des données



# Annexe

## Data Set Information

This data set contains brain signals from three subjects, as well as the time courses of the flexion of each of five fingers. The task in this competition is to use the provided flexion information in order to predict finger flexion for a provided test set.

## Attribute Information

The goal of this element of the competition is to predict the flexion of individual fingers from signals recorded from the surface of the brain (electrocorticography (ECOG)).

Signals from the electrode grid were amplified and digitized using Synamps2 amplifiers (Neuroscan, El Paso, TX). The general-purpose BCI system BCI2000 [1] provided visual stimuli to the patient, acquired brain signals from the Synamps2 system, and also recorded the flexion of individual fingers (on the hand contralateral to the implanted grid) using a data glove (Fifth Dimension Technologies, Irvine, CA). BCI2000 stored the brain signals, the timing of stimulus presentation, and the flexion of each of the fingers in a data file. Data files were converted to Matlab format for this competition. Each patient had subdural electrode arrays (Ad-Tech, Racine, WI) implanted. Each array contained 48-64 platinum electrodes that were configured in 8x6 or 8x8 arrangements. The electrodes had a diameter of 4 mm (2.3mm exposed), 1 cm inter-electrode distance, and were embedded in silastic. Electrocorticographic (ECoG) signals (i.e., 62, 48, and 64 channels from subjects 1, 2, and 3, respectively) were acquired with respect to a scalp reference and ground (fig 1), band pass filtered between 0.15 to 200 Hz, and sampled at 1000 Hz.

