

Exercice 1**8 points**

Vous venez d'arriver en stage, votre patron vous demande de l'aider à définir la stratégie de sa start up par rapport au problème de sélection de variables dans leur super module existant d'apprentissage statistique

1. Qu'est-ce que la sélection de variable et à quoi ça sert ?
2. Quelles sont les méthodes de sélection de variables les plus importantes et comment fonctionnent-elles ?
3. Comment faire pour mettre en œuvre une méthode de sélection de variable sur de vraies données ?
4. doit on utiliser une méthode de sélection de variable lorsque l'on utilise des réseaux de neurones profond (deep learning).

Exercice 2**BCI****12 points**

Le but de cet exercice est d'écrire la fonction matlab permettant de prédire l'étiquette (-1 ou 1) une entrée future, à partir d'un vecteur de 22 composants (dont certaines sont inutiles). Cet ensemble de données contient une représentation par 22 variables des signaux EEG de l'image motrice, c'est-à-dire des signaux EEG enregistrés alors que les sujets évoquent des mouvements des membres, ici des mouvements de la main (codés 1) ou du pied (codé -1). Pour ce faire, nous disposons des données présentes dans le fichier `BCIcsp.mat` disponible sur moodle.

- les entrées `xapp` pour l'apprentissage et `xtest` pour le test
- les étiquettes `yapp` pour l'apprentissage et `ytest` pour le test y

```
>> load BCIcsp.mat
>> whos
Name          Size          Bytes   Class   Attributes
data          1x1            479296  struct

patient = 1 ;
xapp = data.train{patient}(:,1:feat);
yapp = data.train{patient}(:,23);
yapp = ((yapp > 1.5) - 0.5)*2;
xtest = data.train{patient}(:,1:feat);
ytest = data.train{patient}(:,23);
ytest = ((ytest > 1.5) - 0.5)*2;
>>
```

La méthodologie est très important, pensez à justifier vos choix, les critères et les procédures de validation que vous mettrez en œuvre.

1. Proposez un modèle parcimonieux permettant de prédire au mieux pour le patient 1.
2. Proposez un modèle parcimonieux permettant de prédire au mieux pour tous les patients séparément

```
for patient = 1:9 ;
    xapp(:, :, patient) = data.train{patient}(:, 1:feat);
    yapp = data.train{patient}(:, 23);
    yapp(:, patient) = ((yapp > 1.5) - 0.5)*2;
    xtest(:, :, patient) = data.train{patient}(:, 1:feat);
    ytest = data.train{patient}(:, 23);
    ytest(:, patient) = ((ytest > 1.5) - 0.5)*2;
end
```

3. Proposez un modèle parcimonieux permettant de prédire au mieux pour tous les patients ensemble
4. Laquelle de ces approches vous semble la plus pertinente