Chaïmaa Kadaoui / Pierre-Alain Langlois

Age prediction from EEG signals

...

A project for the course:
Sparse wavelet representations and classification

Introduction

Goal

Predict the ages of subjects given their brain activity during sleep:

- Electroencephalogram (EEG)
- Hypnogram

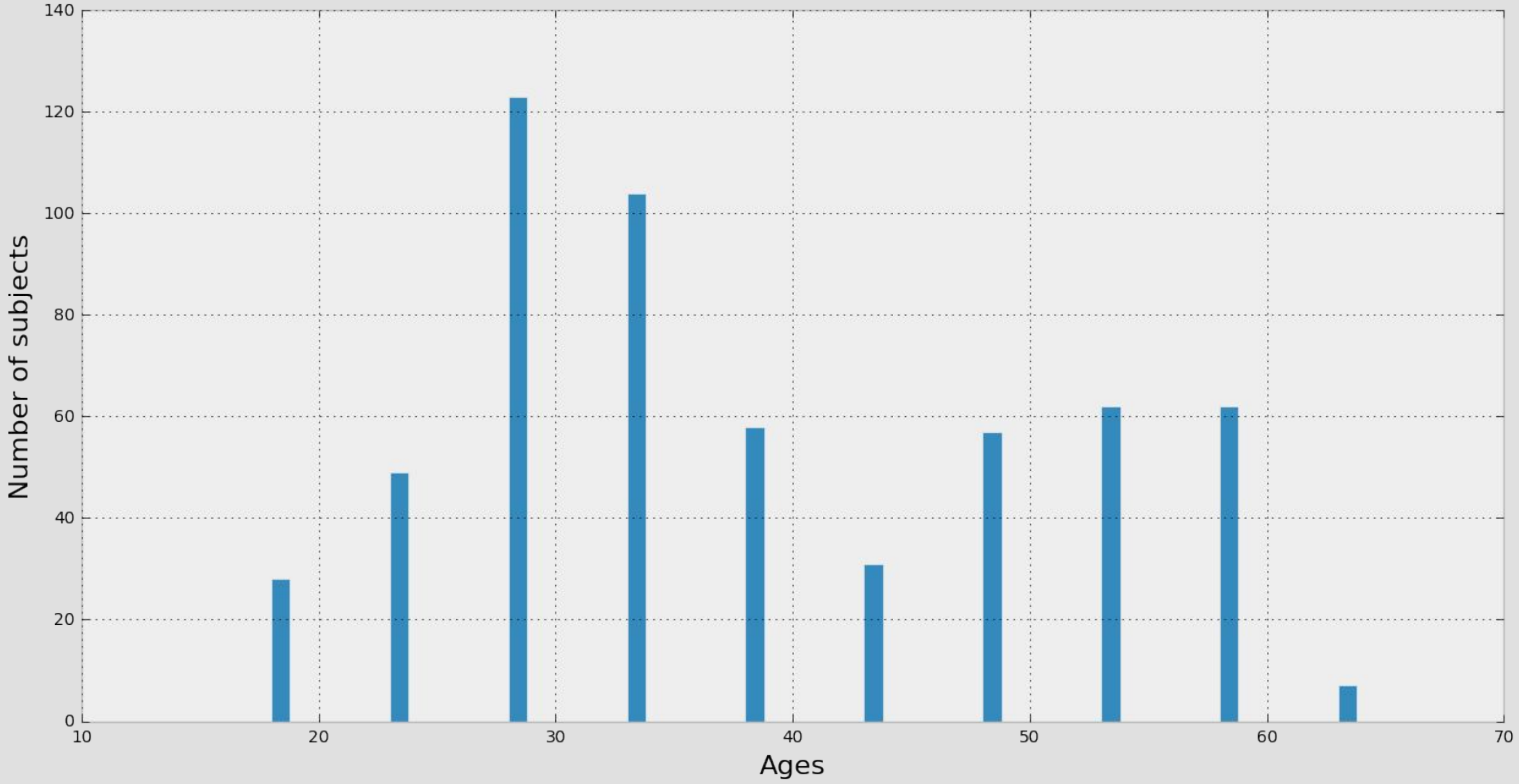
Description of datasets

581 rows for the train set and 249 rows for the test set.

75003 columns each:

- ID: Unique ID for each subject
- DEVICE: The device used to record brain activity (one of 2 kinds)
- EEG: 75000 columns for 5min of the EEG signal during deep sleep (250 Hz sampling rate)
- Hypnogram: A string representing the hypnogram for each subject

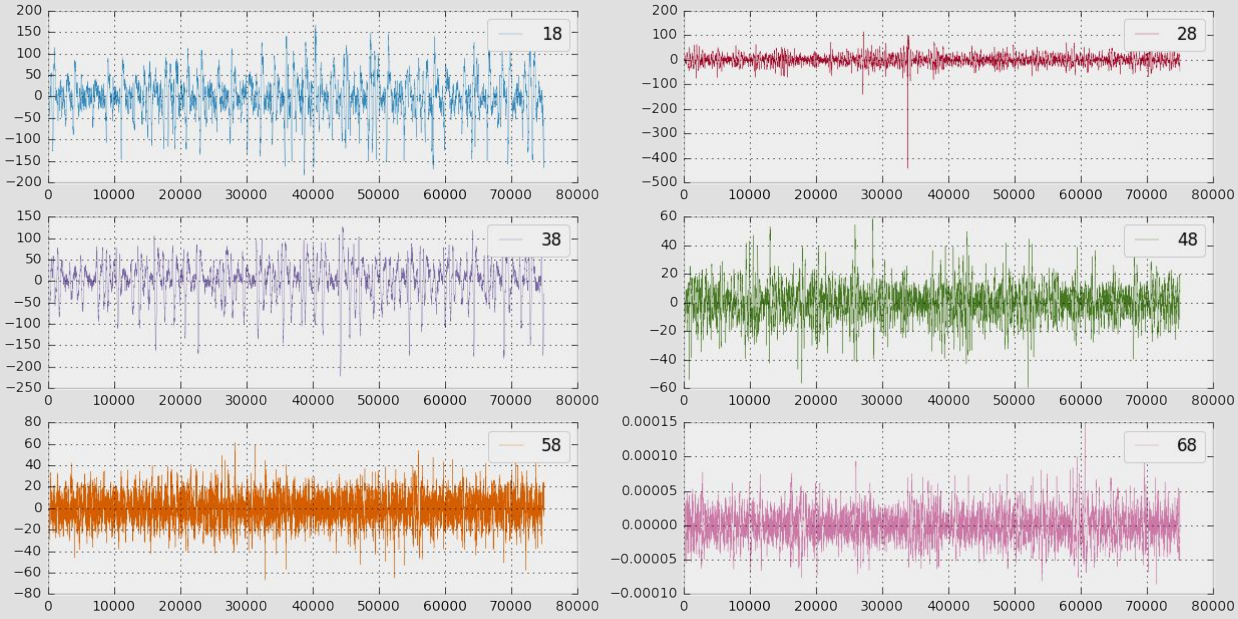
Histogram of ages



Electroencephalogram

Electroencephalogram the data

- Noisy
- No trend
- Not easily characterizable



Examples for a few individuals

First strategy

$$X = TP^T + E$$
$$Y = UQ^T + F$$

- Dimensionality reduction
- Linear solution : PLS
  (Principal Least Square analysis)
- Aim : maximizing covariance between T and U

X,Y : original data ; T,U : new data ; P,Q : projection matrix ; E,F : error

Electroencephalogram

First strategy

- Dimensionality reduction
- Linear solution : PLS
  (Principal Least Square analysis)
- Aim : maximizing covariance between T and U

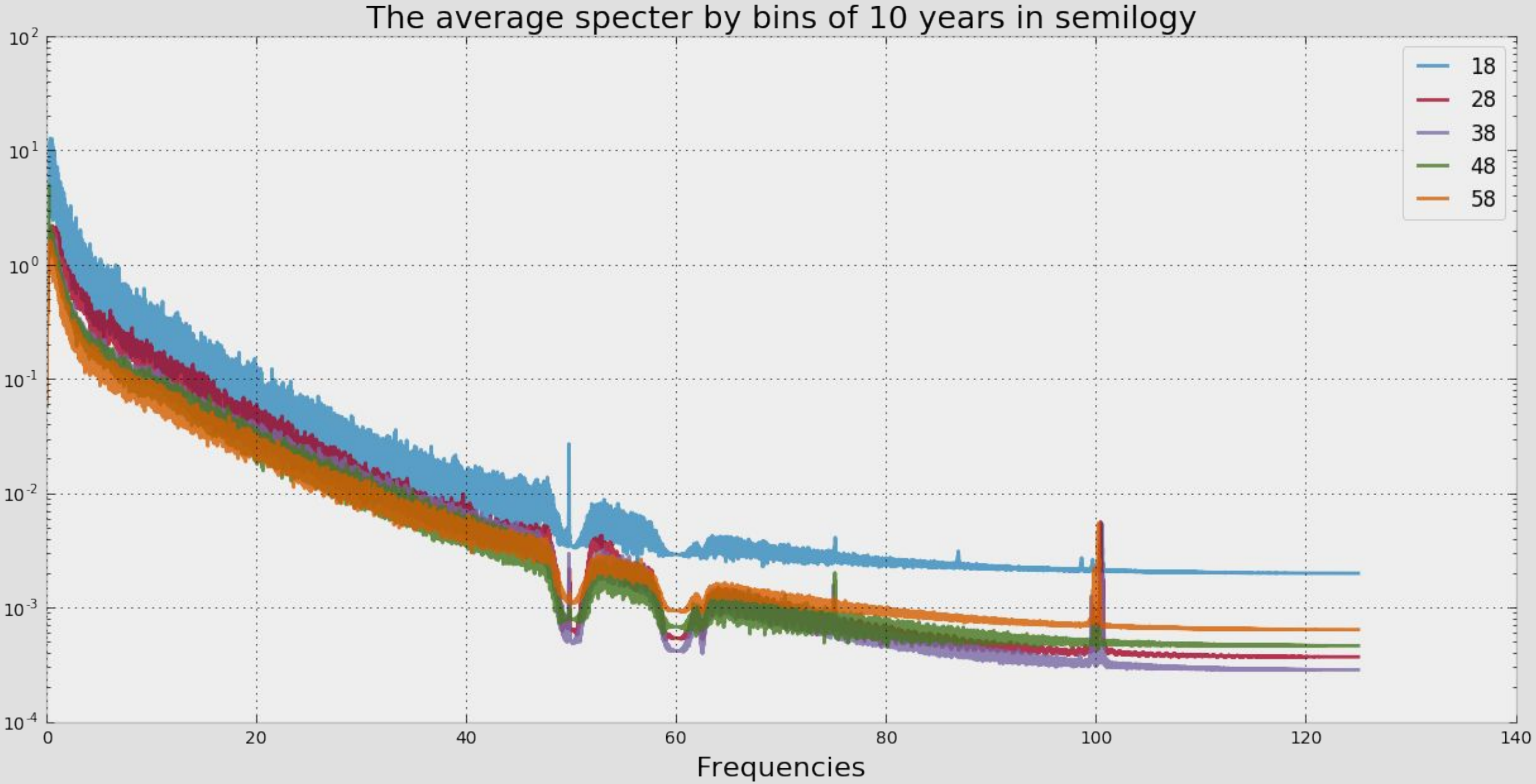
$$X = TP^T + E$$
$$Y = UQ^T + F$$

Same age always returned 41 years old

X,Y : original data ; T,U : new data ; P,Q : projection matrix ; E,F : error

Electroencephalogram

Fourier approach



Electroencephalogram

Neural Network

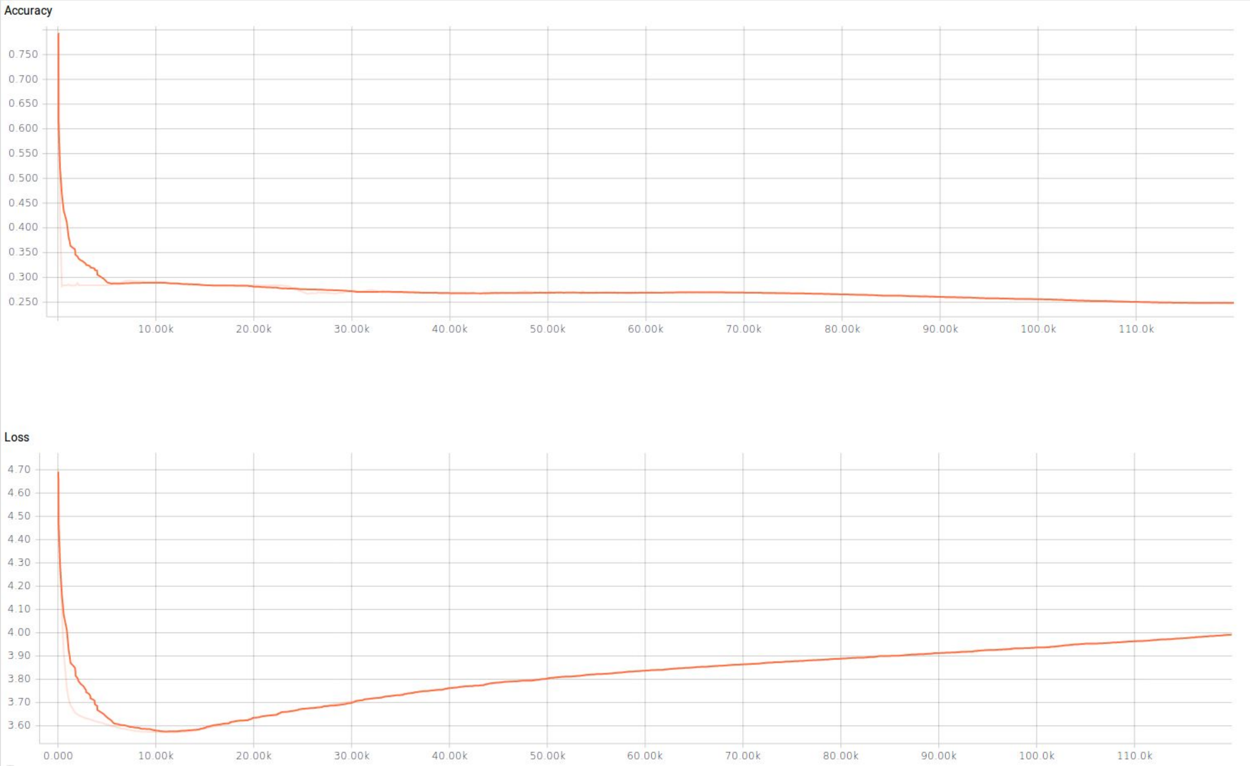
- Lowest frequency : 5 Hz
- Highest frequency : 20 Hz
- Order of the Butterworth filter : 2
- Dimension of the output : 90 (ability to predict age between 0 and 89 years)
- Size of the hidden layers : 200
- Number of layers : 3
- Size of an observation : 5 000 (which means 20s instead of 5 minutes)
- Batch size : 4000
- Non linearity : sigmoid
- Learning rate : 1e-5
- Number of iterations : 100 000
- Dropout probability : 0.5

One more Hypothesis : stationarity

Accuracy : Mean average percentage error

Loss : Softmax cross-entropy with logits

Electroencephalogram Neural Network



Validation accuracy (top) and loss (bottom)

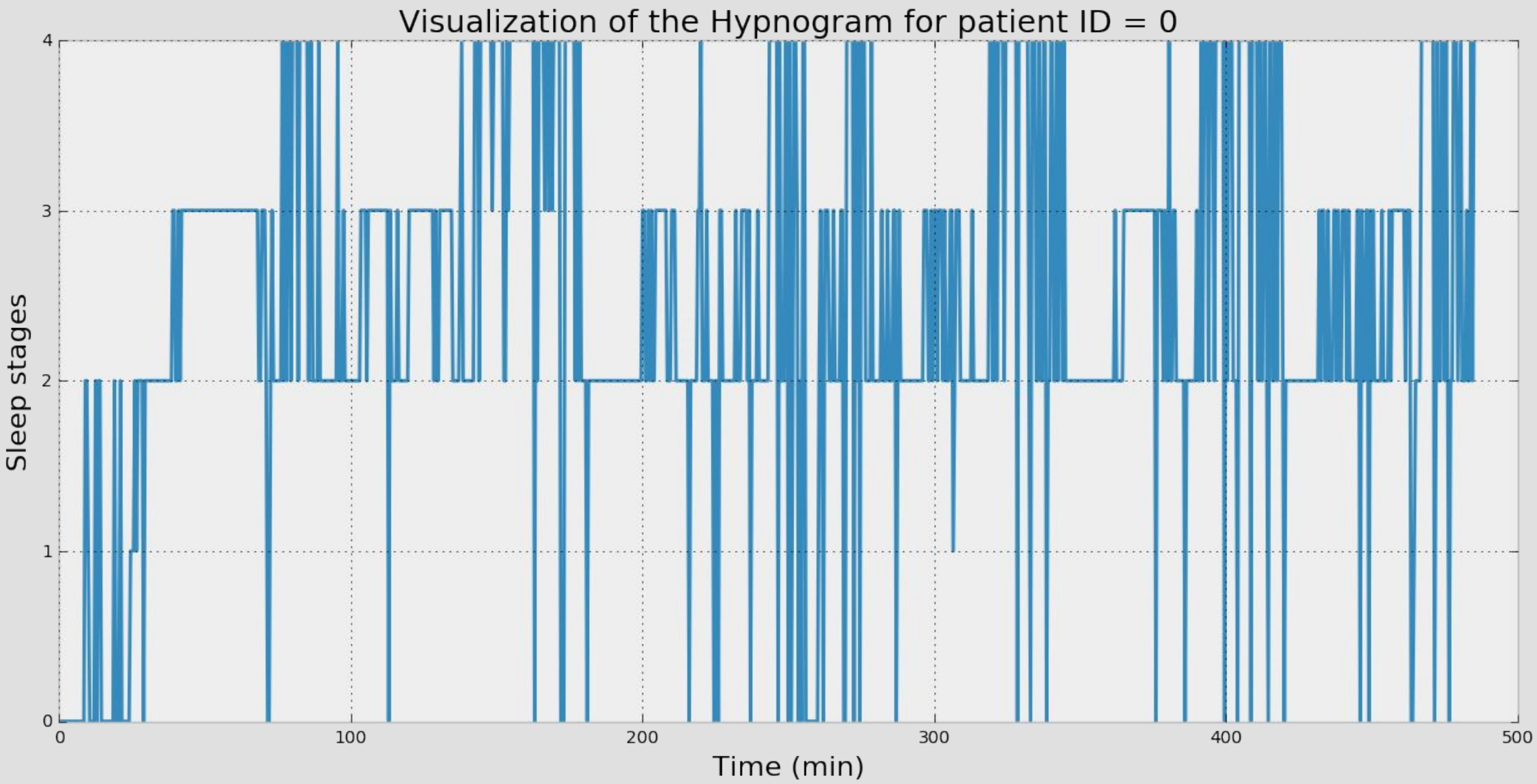
Hypnogram

Hypnogram Sleep stages

The sleep can be decomposed in multiple cycles with different stages:

- Non Rapid Eye Movement
  - N1
  - N2
  - N3
- Rapid Eye Movement
- Awakening

Hypnogram Visualisation



Hypnogram Issues

- The hypnograms are stored as **lists of different sizes** in one column
  => The trivial solution to expand this column into multiple ones won’t work

- The lists contain numerical elements but they represent **categorical data** (1 for Stage 1, 2 for Stage 2,...)
  => Algorithms might induce ordering while this is not the case

- There are **missing values**

➡ Preprocessing this feature is a key

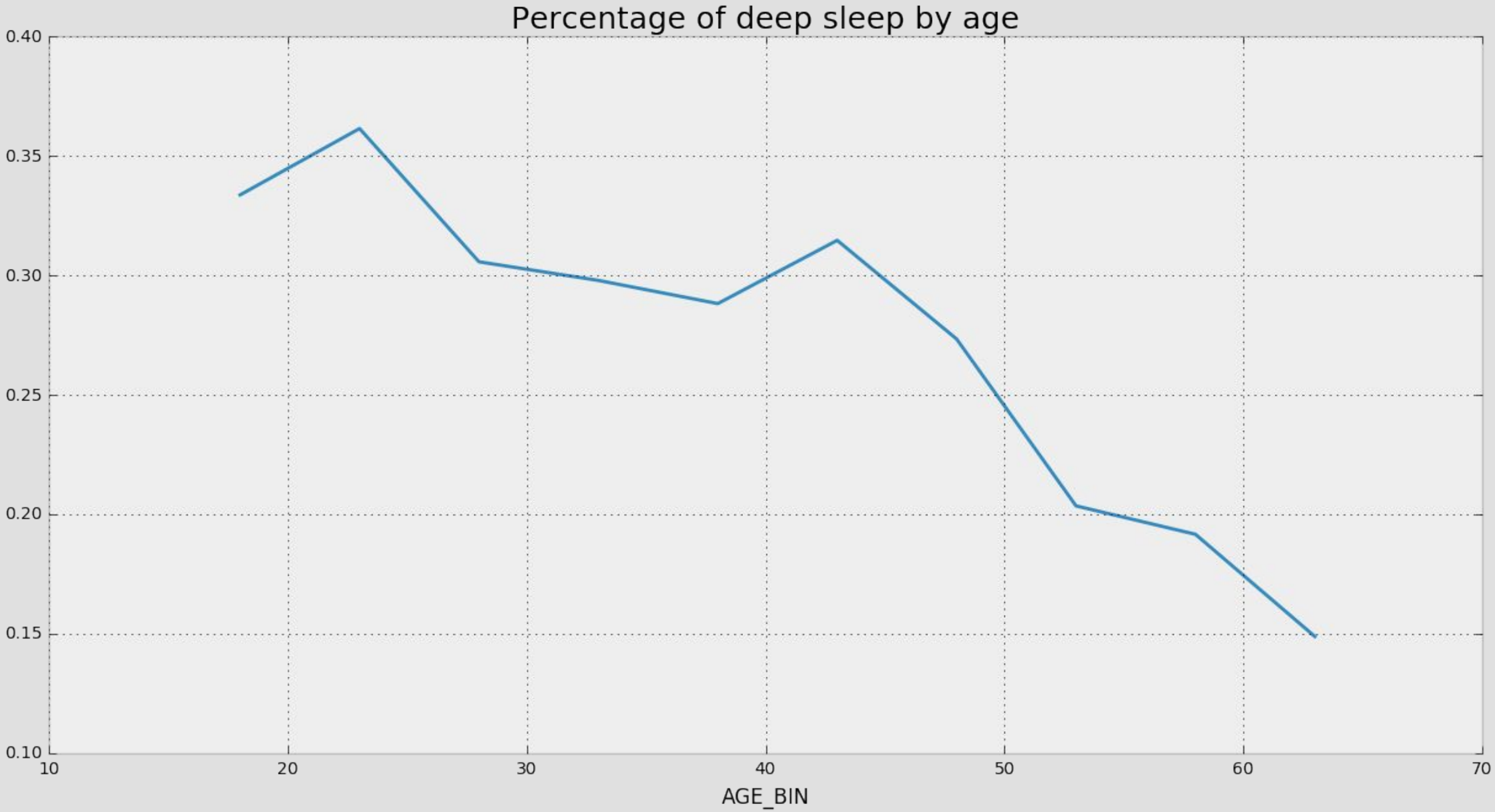
Hypnogram Pre-processing

1. Filling missing values: propagating the last valid value forward

2. Extracting new features thanks to the literature:
    a. Total sleep time (TST)
    b. Sleep efficiency (SE)
    c. Sleep latency (SLAT)
    d. Percentage of each stage including wake time (Si_PERC)
    e. Average duration of each stage (Si_MEAN) and its maximum (Si_MAX)

=> 18 new features

Hypnogram Evolution of deep sleep by age



Hypnogram Random Forest Regressor

Spans multiple decision trees (estimators) and takes the mean prediction

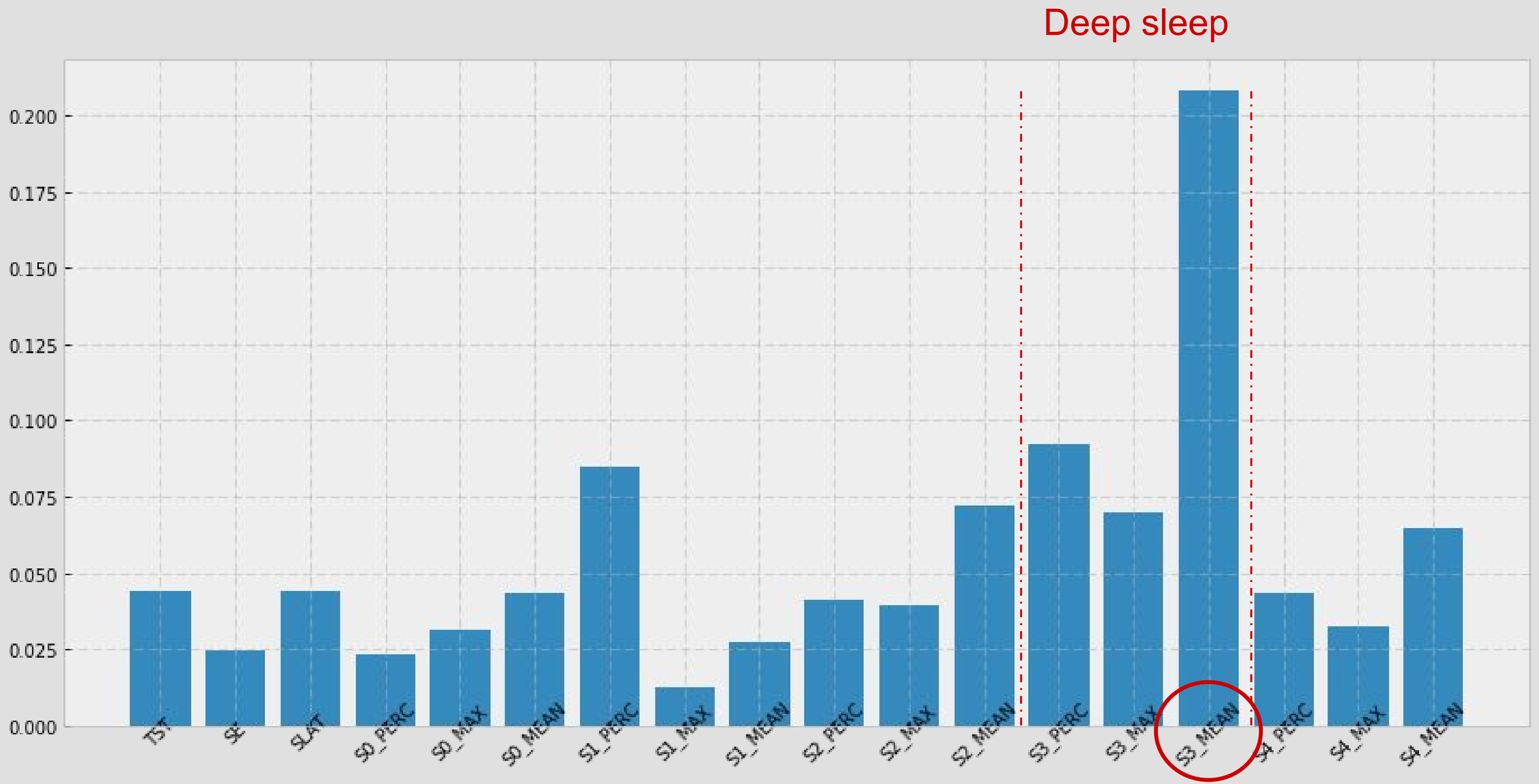
Why ?

- Doesn’t need variables to be of the same scale
- Robust to irrelevant features
- Can estimate feature importance

To avoid over-fitting

- Use a high number of estimators (500)
- Computes the Out-of-bag error (mean prediction error using only the trees that didn’t fit for each sample)

Hypnogram Features importance



Results

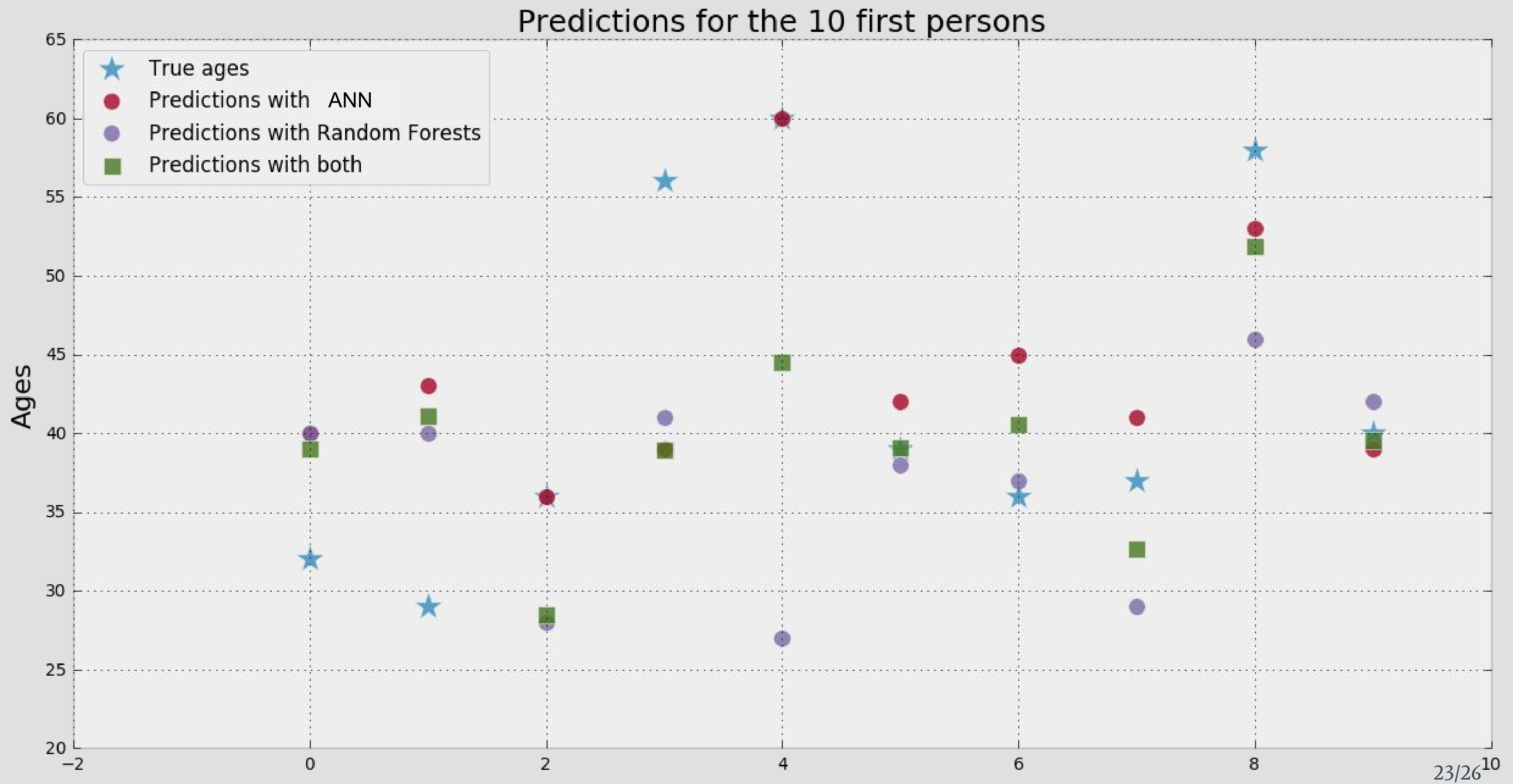
Results

With each approach

- With ANN on the EEG: 21.95 %
- With Random Forest: 21.69 %

After using a Linear Regression on both predictions

- MAPE error: 20.18 %
- Ranking: 11th


Predictions for the 10 first persons
True ages
Predictions with ANN
Predictions with Random Forests
Predictions with both
Ages
65
60
55
50
45
40
35
30
25
20
−2
0
2
4
6
8
10

Discussion

- Getting used to the Tensorflow framework
- Importance of pre-processing: extract meaningful information to avoid non relevant solutions
- More context would be helpful:
  - Influence of external parameters (weather, smoking...)
  - Information about the devices used

Thank you

References

1. D. Dijk, J. Groeger, N. Stanley, and S. Deacon. Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep. 2010.
2. H. Landolt, D. Dijk, P. Achermann, and A. Borbély. Effect of age on the sleep eeg: slow-wave activity and spindle frequency activity in young and middle-aged men. 1996.
3. L. Novelli, F. Raffaele, and B. Oliviero. Sleep classification according to aasm and rechtschaffen and kales: effects on sleep scoring parameters of children and adolescents. 2009.
4. M. Ohayon, M. Carskadon, C. Guilleminault, and M. Vitiello. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. 2004.
5. E. Tagluk, N. Sezgin, and M. Akin. Estimation of sleep stages by an artificial neural network employing eeg, emg and eog. 2009.