EEG Naïve Bayes Classifier (NBC) for Image Induced Brain Activity. Part 2. Analysis in Frequency Domain

The training data set is defined over Cartesian product $S = \{L \times M \times N\}$ of L time samples of electrical potentials measured by N electrodes for M testers. The set of experimental data [1] consists of electrode potentials

$E_k^j(t_i)$: $i \in [1..L]; j \in [1..M]; k \in [1..N]$.

In these experiments, $L = 256,\ M = 17,\ N = 25$. The duration of each observation was $T = 1\ sec,$ thus $t_i = i/L;\ i \in [1..L]$.

The method of Naïve Bayesian Classification considered in Part 1 used all data from S collected for two labels *F(lower)* and *G(rey)*.

The NBC principles

In Part 1 of this analysis (see the previously posted document, *BrainEmotion – NaiveBayesClassifier*), the probability distributions of the k-th electrode's potentials $E_k^j(t_i)$ considered as random variables defined on the subsets of S: $S_k = \{L \times M\} \subset S$ were calculated for each electrode as conditional probabilities given the label:

$$P(E_1|F), P(E_2|F), \dots, P(E_N|F).$$

$$P(E_1|G), P(E_2|G), \dots, P(E_N|G).$$

(The meaning: the probability of potentials for each electrode and each label are calculated over all collected *L* temporal samplings for all testers put together.)

These calculations amount to the process of **training**.

To **classify** a given vector $E^{new} = (E_1^{new}, E_2^{new}, \dots, E_N^{new})$ the following formula is applied:

$$Y^{new} \leftarrow \arg\max_{y_r} P(Y = y_r) \prod_k P(\,E_k^{new} | Y = y_r)$$

Here $Y[r = 1, r = 2] = (y_1 = Flower, y_2 = Grey)$ is the vector of classes and Y^{new} is either $y_1 = Flower$ or $y_2 = Grey$ depending on which of the two products are greater.

Note. The multiplicative criterion used above can be replaced by an additive monotonous and convex function to improve computational stability. The log function being applied leads to the following additive criterion:

$$Y^{new} \leftarrow \arg\max_{y_r} P(Y = y_r) \sum_k \log P(\,E_k^{new} | Y = y_r)$$

NBC in frequency domain

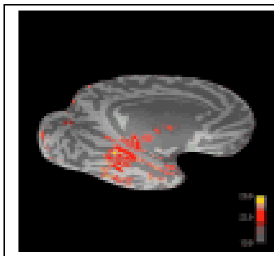
In *Part 2* presented here, the temporal analysis is replaced with consideration of the frequency domain.

The reason for this is the fact that frequency bands have always been considered as very distinctive features in analysis of cognitive and emotional brain activities.

Especially **theta** waves have been linked to experiencing **emotions**, daydreaming, intuition, relaxation, the subconscious mind, etc.

Here is the list of frequency bands commonly used:

- Delta: 1.5-6 Hz - deep, dreamless sleep and in very deep, meditation where awareness is fully detached
- Theta: 6.5-8.0 Hz – emotional response (see above)
- Alpha – deep relaxation (usually when the eyes are closed, when slipping, or during light meditation, intuition
    - Alpha1: 8.5-10 Hz
    - Alpha2: 10.5-12.0 Hz
- Beta – heightened state of alertness, logic and critical reasoning
    - Beta1: 12.5-18.0 Hz
    - Beta2: 18.5-21.0 Hz
    - Beta3: 21.5-30.0 Hz
- Gamma: above 40 Hz – bursts of insight and high-level information processing.



Theta Inflated cortex with EEG activity in the medial temporal lobe during a visually presented semantic memory task. The activity represents the spatiotemporal map of the theta-band activity (~5Hz) which is known to modulate recall and encoding in the medial temporal lobe structures.

The relationship between theta and gamma band activity during cognitive and emotional processing has been recently studied.

Algorithm

Step 1. Apply Fourier Transform $\mathcal{F}(x(t))$ to time dependent functions $E_k^j(t_i)$ for each electrode and each tester to obtain the frequency dependent functions

$$\overline{\boldsymbol{E}}_k^j(f) = \mathcal{F}(E_k^j(t_i))$$

Step 2. Divide the frequency domain of $\overline{\boldsymbol{E}}_k^j(f_i)$ onto the ranges of frequency bands as described above to obtain the spectral densities, $\delta, \theta, \alpha, \beta, \gamma$, of the signal for each frequency band:

$$\delta = \frac{1}{4.5}\int_{1.5}^{6} [\overline{\boldsymbol{E}}_k^j(f)]^2 df$$

$$\theta = \frac{1}{1.5}\int_{6.5}^{8} [\overline{\boldsymbol{E}}_k^j(f)]^2 df$$

$$\alpha = \frac{1}{3.5} \int_{8.5}^{12} [\overline{\boldsymbol{E}}_k^j(f)]^2 df$$

$$\beta = \frac{1}{17.5} \int_{12.5}^{30} [\overline{\boldsymbol{E}}_k^j(f)]^2 df$$

$$\gamma = \frac{1}{20} \int_{40}^{60} [\overline{\boldsymbol{E}}_k^j(f)]^2 df$$

Step 3. Random variables $\delta, \theta, \alpha, \beta, \gamma$ are considered on the sets $\{S_\delta,\ S_\theta,\ S_\alpha,\ S_\beta,\ S_\gamma\} = \{N \times M\}$.

The following conditional probabilities given the label are to be calculated (the training phase):

$$P(\delta|F), P(\theta|F), P(\alpha\ |F), P(\beta|F), P(\gamma\ |F).$$

$$P(\delta|G), P(\theta|G), P(\alpha\ |G), P(\beta|G), P(\gamma\ |G).$$

(The meaning: the probability of spectral densities are calculated over all collected temporal samplings for all testers and all electrode put together.)

These calculations amount to the process of **training**.

To **classify** a given 5-vector $X^{new} = (\delta^{new}, \theta^{new}, \alpha^{new}, \beta^{new}, \gamma^{new})$ the following formula is applied:

$$Y^{new} \leftarrow \arg\max_{y_r} P(Y = y_r) \prod_k P(\,X_k^{new}|Y = y_r)$$

Here $Y[r = 1, r = 2] = (y_1 = Flower, y_2 = Grey)$ is the vector of classes and Y^{new} is either $y_1 = Flower$ or $y_2 = Grey$ depending on which of the two products are greater.

Note. The multiplicative criterion used above can be replaced by an additive monotonous and convex function to improve computational stability. The log function being applied leads to the following additive criterion:

$$Y^{new} \leftarrow \arg\max_{y_r} P(Y = y_r) \sum_k \log P(\,E_k^{new}|Y = y_r)$$

FFT with Excel

With the Data Analysis add-on installed, it is possible to perform Fast Fourier Analysis for given set of time samples of waveform. The number of samples should be power of 2. In our case it is 256.

- First the labels should be created: "A1:Time, B1:data (for example, FP1 for the first electrode potentials) , C1:FFT frequency, D1:FFT magnitude, E1:FFT complex (number).
- The waveform *time* column A consists of 256 cells with time in seconds: 1/256, 2/256, 3/256, …, 1.0
- 256 potentials of a given waveform for a specific electrode are copied to column B,.

- In the "Data Analysis" option, *Fourier Analysis* should be selected.
- The "Input Range" field should contain the waveform data (column B).
- Select "Output range" and set it to the first empty cell of the column were the FFT result shall be presented (D2 in our example).

- The FFT results appear in column D in complex number fom.
- In column E2 calculate the amplitude using expression "=IMABS(D2".
- In column C the frequency step should be calculated from 1 to 128 Hz (the first half of the output data).
- The FFT spectrum is to be plotted as a function of frequency.
- The simple way to calculate the spectral power for all 5 frequency bands will be just find the mean value for each of the given ranges: DELTA=SUM(D2:D8)/7, THETA=SUM(D9:D10)/2, etc.

Algorithm implementation

The goal of this analysis was to show the steps needed to perform the NBC algorithm – not the analysis of the given dataset itself.

The Excel file, FFT Analysis.xlsx, consists of four spreadsheets: Flower Electrode Potentials and Grey Electrode Potentials where the given datasets for one of the testers are presented; as well as FFT for FP1 Flower, and FFT for FP1 Grey which are analysis results given just for one electrode, FP1.

In any case, the means to be used should be more specialized and powerful unlike the Excel program. Thus, the presented spreadsheet is just for illustration of the principals.