**Neurocomputing**
709.014 22S

# Homework assignment 1

**Student:** **Alessio Tonello** reg. number: 12141042

The goal of the seminar is to understand how it is possible to quantify couplings between EEG time-series. The quantification can be performed in both time and frequency domain.

The easier technique in the time domain is to evaluate the correlation between time-series. Doing that, you are understanding if two time-series are connected; and so, if the brain regions that produce those time-series are working together. This is called functional connectivity. However, one important information missed is the causality of this connection, so which time series is influencing the other. If you understand that, you can better understand the flow of information between different brain regions. This is called effective connectivity.

In time domain, a study of effective connectivity of time-series is based on an important class of linear time-independent discrete models called multivariate autoregressive models (MVAR). Those models are a generalization of the simpler autoregressive model (AR).

In this assignment, I will implement from scratch a MVAR model, fitting it on a 9 channels acquisition of 5 seconds with sampling frequency of 128 Hz. The model will try to predict the value of all channels at a time point N, based on a linear combination of the p previous samples of the that channel and all the others. Changing the value of p will influence the prediction of the model. As a result, I will select the value of p (that is called model order) that leads to the best prediction by applying two different techniques of model selection: one based on a two-fold cross validation, and one based on statistical criteria like AIC and BIC.

## Model order selection based on two-fold cross validation (CV)

As suggested in the assignment, I divide the dataset in two folds. One will be the training set and the other the test set. So, on the first one I estimated the parameters of the model and in the other one I will compare the prediction of the model with the real time series. A variable J, which can be seen as a cost function, will represent how good is the model to predict the time-series. The next step is to repeat the procedure be switching the 2 folds. The final cost function value will be the mean of the 2 values of J derived from each fold.
The two-fold CV will be performed for different values of $p = 1,2, \dots p_{max}$; obtaining J(p). The best value of p will be that one that minimize J.
The value of the obtained cost function can be seen in fig.1. The minimum value of J(p) is for p = 3.

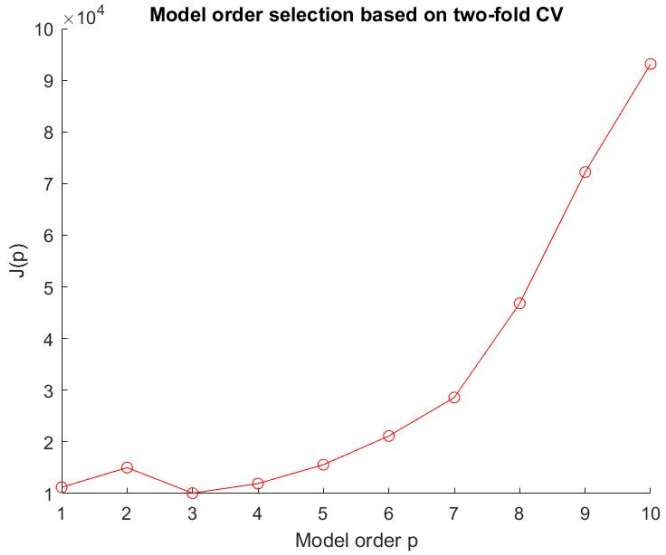


*Figure 1.*

**Model order selection based on AIC and BIC**

This approach is a little bit easier respect to the first one since I will only fit the model on the whole dataset and evaluate the fitting by calculating the value of AIC and BIC. This will be performed for $p = 1,2, \dots p_{max}$ and the best value of p will be that one that minimize AIC(p) and BIC(p). However, usually these criteria tend to overestimated the optimal p value. As a results, if the value of AIC and BIC is similar between several values of p, is always a good idea to select the simpler model, so the smallest value of p. Graphically, this will result in finding the knee of the curve.
The results of this model order selection technique are displayed in fig.2. The best model order value for AIC is p=3, for BIC is p=1. This discrepancy does not surprise me, since the two best values are close.

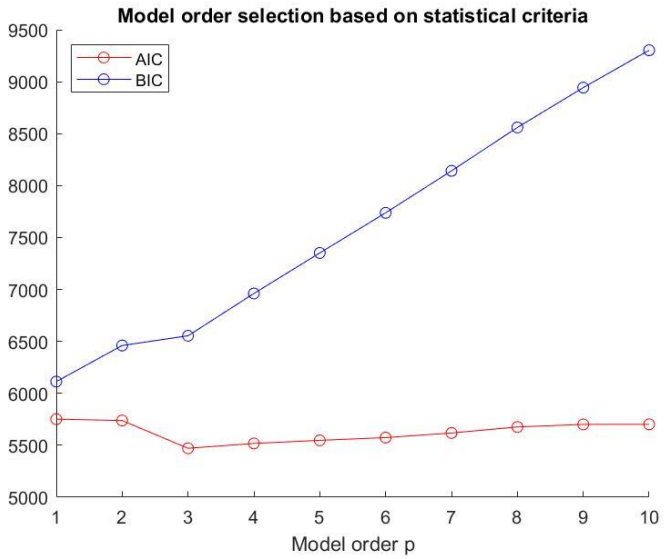


*Figure 2.*

To sum up, by evaluating both techniques, I will select as best model order p = 3. So, I will fit a MVAR of order 3 to the whole dataset, to better catch the lowest frequencies present in the time-series, and to exploit the possible information coming from all the time points.