EE512 – Applied Biomedical Signal Processing

Practical session – Linear Models I

Instructions

- Please submit your report as a single PDF file.
- We recommend working in a group of 3–4 students; you can prepare one single report for the group (`name1_name2_name3_lab_LinearModelsI.pdf`), but every member needs to upload the same file individually.
- Include in the report the code you used when addressing the questions, together with your answers.
- The session is prepared to be done in Matlab – other languages like Python would be possible, but would require much more coding.
- We recommend creating a Matlab script – see “Important recommendation” below.

Useful commands

Type `help function_name` for more information on inputs and outputs.

`xcorr`	Estimate auto and cross-correlation
`aryule`	AR model estimation, autocorrelation method; uses Levinson-Durbin
`arcov`	AR model estimation, covariance method
`a = X \ c`	Solves equation `Xa = c` to estimate `a` in the least-squares sense
`AR2`	Generate a signal using an AR(2) process
`AR_order`	Estimate optimal model order using MDL (and `arcov` for estimation)
`test_white`	Whiteness test

`figure, plot`	Plot signals for visualization. The following example code illustrates how it could be used for a given signal `x` in time:
	`n = size(x,1);`
	`fs = 100; % Sampling frequency in Hz`
	`t = (0:n-1)'/fs; % Sampling times in sec`
	`figure('Units','centimeters','Position',[0,0,17,17],'Color','w');`
	`subplot(3,1,1); plot(t,x); title('Signal'); xlabel('Time (s)');`
	`subplot(3,1,2); etc…`

Important recommendation: Write the commands of your analysis in a Matlab script, starting with a `clear all` command so that your functions start on an empty Matlab workspace. This helps keeping track of what you are doing and can avoid getting confusing results when variables are changed by accident with repeated commands (which can often happen in the command window). At the end, you can simply copy the relevant code from the script to your report.

Experiment 1: comparing AR estimation methods

The function `AR2` generates an AR(2) process with parameters $a_1 = -0.5562$ and $a_2 = 0.8100$.

a) Use `AR2` to generate a signal with 100 samples. Plot the resulting signal.

b) For the generated signal, estimate the model parameters by (i) using the `aryule` function, and (ii) linear prediction (MSE criterion) via directly solving the Yule-Walker equations: `a = -R\b`, where:

- `R = [Rxx(100) Rxx(101); Rxx(101) Rxx(100)]`
- `b = [Rxx(101); Rxx(102)]`
- `Rxx` are the biased estimates of the autocorrelation (`xcorr(x,'biased')`).

Do you get similar or different results? Why was this to be expected? (*note: remember you can use* `help aryule` *for more info*)

c) For the generated signal, estimate the model parameters by (i) using the `arcov` function, and (ii) linear prediction (MSE criterion) using the covariance approach: `a = -ϕ\b` where:

- `ϕ = [ph(1,1) ph(1,2); ph(2,1) ph(2,2)]`
- `b = [ph(1,0); ph(2,0)]`
- `ph = @(j,i) sum(x(1+2-j:n-j).*x(1+2-i:n-i))`

Do you get similar or different results? Why was this to be expected? (*note: remember you can use* `help arcov` *for more info*)

d) Try estimating the AR parameters using both the autocorrelation approach (`aryule`) and covariance approach (`arcov`), for different signal sizes: 10, 100, 1000. Do the estimates of the two approaches become closer or more distinct with increasing signal size? Why?

Experiment 2: classifying EEG signals

The file `bci.mat` contains two data matrices, `left_hand` and `right_foot`, from a brain-computer interface (BCI) experiment. Each column in these matrices corresponds to a 2-second EEG recording (sampling frequency of 128 Hz) from the same electrode. The recordings in `left_hand` (respectively `right_foot`) were performed while the subject imagines a movement of the left hand (resp. right foot). The goal of the BCI experiment is to be able to "guess" what is being imagined based on the EEG signals alone.

a) Estimate the AR order for each signal of each matrix (the plots from `AR_order` are not required). Is there already a difference between the two categories overall? And if we wish to perform AR model estimation using a common choice of model order for all signals, which value should be chosen, and why?

b) Estimate the AR models using the autocorrelation method (`aryule`) with order 3. On which coefficients is the separation between categories most promising?

Experiment 3: AR model evolution over time

Real-life physiological signals can often vary substantially (and meaningfully) throughout a recording. It's usually good practice to have a look before trying to apply a model. Consider the signal in `AF_sync.dat` – a recording of ECG atrial activity during atrial fibrillation (sampling frequency of 50 Hz).

a) Plot the signal and consider its evolution over the course of the recording. At the start (until sample ~2000 approximately), the signal is moderately organized; then it becomes very organized until sample ~3000. This probably corresponds to a drastic reduction in the number of fibrillatory waves in the atrial tissue (flutter). In the last part of the recording, the fibrillation, and thus the signal, becomes very disorganized. Identify and mark these three distinct periods in your plot (or equivalently, present zoomed windows of each).

b) Consider a segmentation of the signal into 500-sample windows with 50% overlap. For each segment, estimate (i) the signal variance, (ii) optimal AR order, and (iii) the AR coefficients & excitation variance (use `arcov`). Plot the time evolution of the raw signal, the AR order, the signal variance, the excitation variance, and finally the ratio of excitation variance to signal variance. Interpret their time evolution, and how they relate to the organization of the signal in the three afore-mentioned stages.

   Hint for the segmentation: the following code snippet defines the segmentation windows and extracts a feature from each segment (in this example: the standard deviation). A similar approach can be used to perform the required estimations.

```
nt = length(AF_sync);        % Signal length
nw = 500;                    % Window length
nv = round(nw * 0.50);       % Overlap length

Ki = []; % Starting index for each segment
Kc = []; % Central index for each segment (useful for plotting)
Kf = []; % Ending index for each segment

% Define the segment indices
ki = 1; kf = nw;
while true
    Ki = [Ki; ki];
    Kc = [Kc; round(0.5*(ki+kf))];
    Kf = [Kf; kf];

    ki = ki+nw-nv;
    kf = kf+nw-nv;

    if kf > nt, break; end
end

% Estimate the STD per segment
nc = size(Ki,1);
sdev = zeros(nc,1);
for kc = 1:nc
    sdev(kc) = std(AF_sync(Ki(kc):Kf(kc)));
end
```

c) Consider the three periods described in a). For each of them, estimate the AR parameters, and then obtain the associated filter poles using the `roots` function. Plot the poles on top of the unit circle (see code snippet below) for each of the three periods. Comment on how signal organization relates to pole location (proximity to the circle).

   Hint for plotting the unit circle:

```
t = linspace(0,2*pi,1000);
plot(cos(t),sin(t));
axis('square');
set(gca,'Xlim',[-1.5 1.5]);
set(gca,'Ylim',[-1.5 1.5]);
```

You can then use `hold on` / `hold off` to add your pole plots.

Experiment 4: recovering the excitation (whitening filter)

Consider the signal in file `speech.dat`, which corresponds to the sound /a/ sampled at 8 kHz.

a) Remove the mean value, and then plot the signal.

b) Determine the optimal model order for this (de-meaned) signal (`AR_order`), and estimate the AR model parameters for that order using `aryule`.

c) Compute the underlying excitation signal, using the command `exc = filter(a,1,x)`, where `x` is the speech signal and `a` is the vector of AR parameters. Explain why the excitation can be estimated in this way!

d) Visualize the excitation signal, and compare it to the output (speech) signal. Does it look like white noise? Test whether it is so using `test_white`.

e) Repeat the same procedure (a–d) to the signal in the 1st column of the file `blood.dat` (daily systolic pressure recorded on a patient). How does the excitation signal compare to that of the speech example?

Have a good session, and don't hesitate to ask questions!