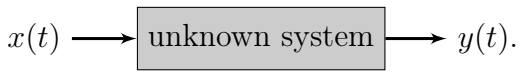
Moving average model with n delays

We have an unknown system, which to an input signal $x(t)$ responds with output signal $y(t)$:



Based on observations, say $N+1$ measurements, we would like to predict the behaviour of the system. Suppose that measurements of the signals $x(t)$ and $y(t)$ are taken at times $t = 0, 1, \ldots, N$. According to our model, the output $y(t)$ at time t can be expressed as a linear combination of n inputs $x(t), x(t-1), \ldots, x(t-n+1)$ from the time interval $[t-n+1, t]$, i.e.

$$y(t) = h_1 x(t-n+1) + h_2 x(t-n+2) + \cdots + h_{n-1} x(t-1) + h_n x(t). \tag{1}$$

To be more precise: Our data consists of $N+1$ inputs

$$x_0, x_1, \ldots, x_N$$

and $N+1$ corresponding outputs

$$y_0, y_1, \ldots, y_N,$$

which are taken at integer times t from the interval $[0, N]$. (With this notation we have $x_i = x(i)$ and $y_i = y(i)$.) Equation (1) becomes

$$y_k = h_1 x_{k-n+1} + h_2 x_{k-n+2} + \cdots + h_{n-1} x_{k-1} + h_n x_k,$$

and we have such an equation for all $k = n-1, n, \ldots, N$. Thus, we have a system of $N-n+1$ linear equations with n unknowns $h_1, \ldots, h_n$ where we assume that $N-n+1 \geq n$. (Actually, we assume that $N-n+1$ is much larger than n.) Let A be the matrix of this system, $\mathbf{h} = [h_1 \ldots h_n]^\mathsf{T}$ the vector of unknowns, and $\mathbf{y}$ the right-hand side. We are looking for the linear least squares solution of the system $A\mathbf{h} = \mathbf{y}$.

Task

1. Derive and write down/describe the general form of the matrix A and right-hand side $\mathbf{y}$. Suppose $\mathbf{h}$ is known. How would you predict the output $y_0, \ldots, y_m$ of the system given only input data $x_0, \ldots, x_m$? Which components of the output can be predicted?

2. Program a Julia function `movavg(x, y, n)` which, given the input data `x = [x0, ..., xN]` and output data `y = [y0, ..., yN]`, finds the coefficients `h = [h1, ..., hn]` for the moving average model with `n` delays. (Follow the specifications: `x`, `y`, and `h` are *row vectors*.)

3. Suppose that all inputs and all outputs are constant, i.e. $x_i = y_i = 1$. First guess and then derive the solution $\mathbf{h}$ for arbitrary N and n. Compare your solution with the output of `movavg`.

4. Program a function `prediction(x, h)` which, given the input data `x` and coefficients `h`, predicts the output `y` using our model.

5. Test your functions on the data sets `io-train.txt` and `io-test.txt` for $n = 1, 2, 3, 5, 10$. (`io-train.txt` is the training set, use it to evaluate $\mathbf{h}$. `io-test.txt` is the test set, test the prediction of the model on this data set.) How does the accuracy of the prediction depend on n?

Submission

Use the online classroom to submit the following:

1. a Julia file **homework1.jl** containing the functions **movavg** and **prediction** which sould be well-commented and contain at least one test,

2. a report file **solution.pdf** which contains the necessary derivations and answers to questions.

While you can discuss solutions of the problems with your colleagues, the programs and report must be your own creation. You can use all Julia functions from problem sessions.