EEL 6504
Due April 25, 2022

Project 2 - Groups of two students

This is a project, so I expect that you work only with your partner in the group. Divide equally the load between the team and state the division of labor. Prepare a 7 pages report in scientific paper format, using the template from IEEE Trans. Signal Processing. The goal is to compare implementations of the QKLMS and QKRLS trained with both MSE and MCC to predict the next sample of the sunspot time series data, which is available in the course website. This data set records the sunspot activity monthly since the XVIII century.

<u>Conventional prediction</u>. Normally, we delay the input to the model by one sample and the current sample will be used for the desired response. Use an input layer with 6 delays to map the data to RKHS (i.e., an input embedding vector of size 6). As a baseline for the comparison, please implement the linear filter with LMS filter trained with the same two cost functions.

Provide an analysis of the three free parameters (kernel size, quantization, learning rate/ forgetting factor) using the prediction error for both systems trained with MSE and MCC in the first 200 samples. Quantify the final prediction error power in the data set, and the tradeoff with computation time. Please create the histogram of the errors in the test set for the two cost functions. Remember to select appropriately the kernel size in MCC (the 4^{th} hyper-parameter) in the training set. Test the performance of the conventional predictor in predicting 1, 10 and 20 samples ahead and summarize in a table the important results. For the comparisons, when using the one sample lag, estimate the prediction error 10 and 20 samples ahead; you can also substitute the single delay between the desired and the input by z^{-10} and z^{-20} (i.e., a delay of 10 and 20 samples), and predict the next sample to verify which method is better.

<u>Generating the trajectory</u>. The final test that you will be doing is to **generate** with the different trained models (LMS, QKLMS/QKRLS) the sunspot time series. To accomplish this, you will have to create a recurrent system i.e., use the trained model and feedback its output to the model input with a delay of one sample. The initialization of this recurrent model must be done carefully to align the inputs and the delayed output. Take 6 samples of the time series and place them in a first in last out buffer (FILO) to create the first input embedding vector to the model; evaluate the model output and append it to the top of the FILO to create the next input to the model; keep iterating this procedure by feeding predicted samples instead of inputs. Since you know the next samples of the time series that follow the 6 samples, you can compute the error between the true next sample and the model output. If the predictions are good, then the model will keep generating samples that look like the sunspot time series. But expect a degradation over time (why?). When the instantaneous deviation from the true sunspot data and the filter output reaches 1/3 of the standard deviation of the sunspot time series you stop and count how many samples the model takes to meet the error threshold. Use 20 different initial conditions and average the prediction steps for reporting.

<u>Extra credit: Learning the trajectory</u>. Iterative prediction trains the model in the same way as it is going to be used for the trajectory generation. Use the filter weights as trained above for sample prediction and the error is still created between the system output and the next sample. But now, keep training the filters with the delayed outputs instead of with the input samples until the error stabilizes at small values. Expect training to be slower than before and reuse the dataset for the desired response if needed. In a table show the number of samples (horizon of predictability) that each of the different trained systems can predict, with an instantaneous normalized error < 0.3. Explain the reason for the results obtained.