# Datamining and Neural Networks: Exercise Session 2

In this session we investigate some examples of time-series prediction and classification using multilayer perceptrons. We make use of the the Neural Network Toolbox in MATLAB version 2010b or later. Consider the following problems:

Feel free to contact us if you still have questions after consulting the documentation and/or your colleagues. Our email addresses are lynn.houthuys@esat.kuleuven.be and joachim.schreurs@esat.kuleuven.be.

## 1 Santa Fe laser data - time-series prediction

The Santa Fe data set is obtained from a chaotic laser which can be described as a nonlinear dynamical system. Given are 1000 training data points. The aim is to predict the next 100 points (it is forbidden to include these points in the training set!). The training data are stored in `lasertrain.dat` and are shown in Figure 1a. The test data are contained in `laserpred.dat` and shown in Figure 1b.

Train a MLP with one hidden layer. Given a certain lag $p$, the training is done in feedforward mode

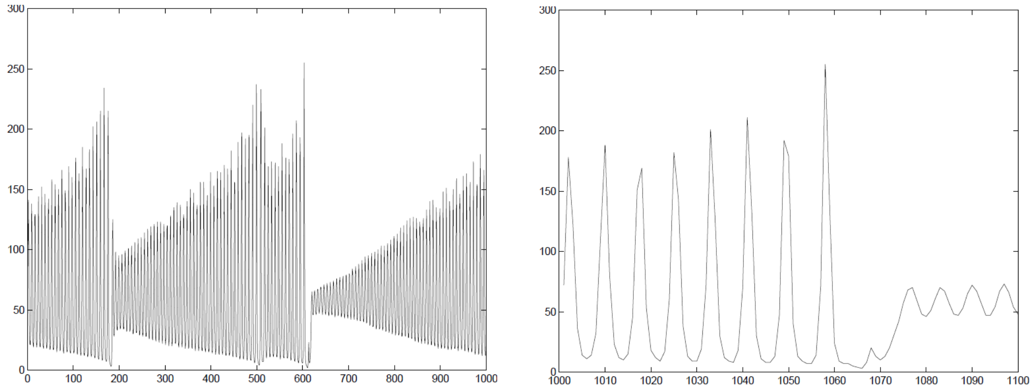
$$\hat{y}_{k+1} = w^T \tanh(V[y_k; y_{k-1}; \ldots; y_{k-p}] + \beta).$$

In order to make predictions, the trained network is used in an iterative way as a recurrent network:

$$\hat{y}_{k+1} = w^T \tanh(V[\hat{y}_k; \hat{y}_{k-1}; \ldots; \hat{y}_{k-p}] + \beta).$$

To format the data you can use the provided function `getTimeSeriesTrainData`. Make sure you understand what the function does by trying it out on a small self-made toy example.

Investigate several methods that you have applied in Exercise Session 1 and compare. Which methods are applicable in which circumstances or not (and why)? Which approach gives the best prediction result? Also try other input vectors for the MLP which do not consist of subsequent points in time: e.g. $[y_k; y_{k-1}; y_{k-2}; y_{k-20}; y_{k-21}; \ldots; y_{k-p}]$ instead of $[y_k; y_{k-1}; y_{k-2}; \ldots; y_{k-p}]$.



(a) Training set

(b) Test set

Figure 1: Training (a) and test (b) as a function of the discrete time index $k$

## 2 Alfabet recognition

Investigate demo `appcr1` [1].

## 3 Breast Cancer Wisconsin - classification problem

Consider the UCI data set of Breast Cancer Wisconsin (Diagnostic) Data Set (file `bcw.mat`). The data set contains 569 records of patients who either have breast cancer (label 1) or not (label 0). Train MLP's for this binary classification problem by taking class labels +1 and -1 as target values in a nonlinear regression, or simply use the build-in matlab routine `patternnet`. Try different numbers of hidden units and compare several methods that you tested in Exercise Session 1 and discuss. It could be good to look at the number of missclassifications as well as the ROC curve.

[1] Make sure you don't have any conflicting toolboxes in your path (like e.g. the LS-SVM toolbox) when running this demo. If you still have troubles running `appcr1`, you can find the complete documentation about the demo here: https://nl.mathworks.com/help/nnet/examples/character-recognition.html