

**COMP3206/6229 (2015/16): Machine Learning Assignment** 20%

Issue	Thursday, 19 November 2015
Deadline	Friday, 11 December 2015 (12:00 Noon)

- This is an assessment and you are expected to work independently.
- Spend no more than 20 hours on this assignment. Parts of it are designed to build on previous short tasks you have done. Please revise those exercises.

## Neural Network Approximation

Artificial neural networks are good function approximators. In the case of pattern recognition, they are used as estimators of posterior probabilities of class membership.

Consider a two-class pattern classification problem in two dimensions, in which each class is Gaussian distributed with distinct means and covariance matrices  $\mathcal{N}(\mathbf{m}_j, \mathbf{C}_j)$ . Take

$$\mathbf{m}_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \quad \mathbf{C}_1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{m}_2 = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{C}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Compute the posterior probability on a regular grid in the input space and plot the decision boundary for which the posterior probability satisfies  $P[\omega_1 | \mathbf{x}] = 0.5$ . Show 100 samples drawn from each of the classes superposed on the same graph. Draw the posterior probability as a three dimensional graph.
- Using the data sampled from each of the distributions, train a feedforward neural network using MATLAB's `Neural Networks` toolbox. The following are commands you are likely to use.

```
[net] = feedforwardnet(20)
[net] = train(net, X, y)
[output] = net(x)
```

Evaluate the network output on a regular grid, and plot the decision contour. How well does it compare with the Bayes' optimal boundary? Compare the approximation at two different sizes of the network.

## Time Series Prediction

Time series prediction is an important application of machine learning techniques. The prediction problem is formulated as a regression problem. Given a series, denoted by  $s(n), n = 1, \dots, N$ , we wish to predict the  $n^{\text{th}}$  sample  $s(n)$  from  $p$  samples that occurred in the past ( $p \ll N$ ):  $[s(n-1) \ s(n-2) \ \dots \ s(n-p)]$ .

### Chaotic Time Series

The Mackey-Glass model [1] is a popular chaotic time series. It is obtained by integrating the nonlinear differential equation,

$$\frac{dx}{dt} = \frac{a x(t-\tau)}{1 + x(t-\tau)^{10}} - b x(t),$$

which, for certain parameter values gives sustained oscillatory signals.

- Obtain the functions `mgfunc.m`, `mackeyglass_rk4.m` and `mackeyglass_eq` from MathWork's contributed software site <http://www.mathworks.co.uk/matlabcentral/fileexchange/24390-mackey-glass-time-series-generator>, and run `mgfunc` to generate a time series of 2000 samples.
- Use the first  $N = 1500$  samples to train a prediction model and the remaining 500 as test data. With  $p = 20$ , construct the design matrix and output of a regression problem. Your input matrix will have  $N - p + 1$  rows and  $p$  columns, with each row being a time shifted version of the previous one.
- Estimate a linear predictor from the training data and check how well it does *one step ahead* prediction on the test data.
- Train a feedforward neural network and evaluate how well it performs on one step ahead prediction.
- Use the trained neural network in a free running mode, feeding back predicted outputs feeding back into the input and check if sustained oscillations are possible [2]<sup>1</sup>.

## Financial Time Series

Obtain daily FTSE100 data for the past five years from a financial data provider *e.g.* `finance.yahoo.com`. Formulate a neural network predictor that predicts tomorrow's FTSE index value from 20 past trading days. Use market `Close` prices. Was there an opportunity to make money using your knowledge of neural networks? Does the ability to predict stock index improve if you were to use past values of `Volume Traded` information as additional input?

## Report

Present your work in a report of no more than six pages.

## Marking Scheme

In each part, 3/5 marks will be awarded for completing the work correctly. 5/5 may be gained for *something extra* in the form of insightful discussion of the problem and its solution.

Neural network approximation	5
Mackey-Glass Time Series	5
Financial Time Series	5
Clarity of presentation	5

## References

- [1] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
- [2] E. A. Wan, "Modeling nonlinear dynamics with neural networks: Examples in time series prediction," in *Proceedings of the Fifth Workshop on Neural Networks*, pp. 327–232, 1993.

---

<sup>1</sup>A copy of [2] is available in the course notes page as `Wan1993.pdf`