Lecture 13. Notes on nonlinear (MLP) models for time series forecasting

Alexey Artemov, Fintech CS HSE

May 29, 2018

1 Introduction: principles of neural modelling

1.1 The basic architecture of multilayer perceptron (MLP)

1.2 An application to time series forecasting: tapped delay neural network (TDNN)

1.2.1 Relation with NARX models

1.2.2 Hybrid ARIMA-MLP models

Original work: (G. P. Zhang 2003).

2 Learning algorithms for MLPs

2.1 Universal approximation property

https://en.wikipedia.org/wiki/Universal_approximation_theorem
Cybenko's work (Cybenko 1989)

2.2 Training MLPs with backpropagation

https://en.wikipedia.org/wiki/Backpropagation

2.3 Training nonlinear regression models: Levenberg–Marquardt algorithm

https://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm
See for advanced general methods: (Fan and Pan 2009; Yoo, Sung, and Lee 2003).
See for single-hidden layer MLPs: (Matias et al. 2014; Li, Shao, and Yiu 2013).
For an affordable application, see (Voyant et al. 2014).

3 Model selection for MLPs

3.1 AIC, BIC and information criteria

Subspace information criterion: (Sugiyama and Ogawa 2001).
Network information criterion: (Murata, Yoshizawa, and Amari 1994).

3.2 Optimal brain damage

Original work on OBD: (LeCun, Denker, and Solla 1990).
Application of OBD to tapped delay neural networks: (Svarer, Hansen, and Larsen 1993).
Something very similar for NARX: (Lin et al. 1997).

3.3 Ensembles of models

For an application to time series analysis, see (Kourentzes, Barrow, and Crone 2014).
(G. P. Zhang and Berardi 2001) provides an application of ensembles for financial time series.

4 Numerical comparisons of MLPs vs. other models

MLPs vs. ARIMA and NARMA: (Allende, Moraga, and Salas 2002).
MLPs vs. FIR and Elman's nets: (Koskela et al. 1996).
Review on forecasting with MLPs: (G. Zhang, Patuwo, and Hu 1998).

References

Allende, Héctor, Claudio Moraga, and Rodrigo Salas (2002). "Artificial neural networks in time series forecasting: a comparative analysis". In: *Kybernetika* 38.6, pp. 685–707.

Cybenko, George (1989). "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4, pp. 303–314.

Fan, Jinyan and Jianyu Pan (2009). "A note on the Levenberg–Marquardt parameter". In: *Applied Mathematics and Computation* 207.2, pp. 351–359.

Koskela, Timo et al. (1996). "Time series prediction with multilayer perceptron, FIR and Elman neural networks". In: *Proceedings of the World Congress on Neural Networks*. Citeseer, pp. 491–496.

Kourentzes, Nikolaos, Devon K Barrow, and Sven F Crone (2014). "Neural network ensemble operators for time series forecasting". In: *Expert Systems with Applications* 41.9, pp. 4235–4244.

LeCun, Yann, John S Denker, and Sara A Solla (1990). "Optimal brain damage". In: *Advances in neural information processing systems*, pp. 598–605.

Li, Leong Kwan, Sally Shao, and Ka-Fai Cedric Yiu (2013). "A new optimization algorithm for single hidden layer feedforward neural networks1". In: *Applied Soft Computing* 13.5, pp. 2857–2862.

Lin, Tsung-Nan et al. (1997). "A delay damage model selection algorithm for NARX neural networks". In: *IEEE Transactions on Signal Processing* 45.11, pp. 2719–2730.

Matias, Tiago et al. (2014). "Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine". In: *Neurocomputing* 129, pp. 428–436.

Murata, Noboru, Shuji Yoshizawa, and Shun-ichi Amari (1994). "Network information criterion-determining the number of hidden units for an artificial neural network model". In: *IEEE Transactions on Neural Networks* 5.6, pp. 865–872.

Sugiyama, Masashi and Hidemitsu Ogawa (2001). "Subspace information criterion for model selection". In: *Neural Computation* 13.8, pp. 1863–1889.

Svarer, Claus, Lars Kai Hansen, and Jan Larsen (1993). "On design and evaluation of tapped-delay neural network architectures". In: *Neural Networks, 1993., IEEE International Conference on*. IEEE, pp. 46–51.

Voyant, Cyril et al. (2014). "Time series modeling with pruned multi-layer perceptron and 2-stage damped least-squares method". In: *Journal of Physics: Conference Series*. Vol. 490. 1. IOP Publishing, p. 012040.

Yoo, Chang Kyoo, Su Whan Sung, and In-Beum Lee (2003). "Generalized damped least squares algorithm". In: *Computers & chemical engineering* 27.3, pp. 423–431.

Zhang, G Peter (2003). "Time series forecasting using a hybrid ARIMA and neural network model". In: *Neurocomputing* 50, pp. 159–175.

Zhang, G Peter and VL Berardi (2001). "Time series forecasting with neural network ensembles: an application for exchange rate prediction". In: *Journal of the Operational Research Society* 52.6, pp. 652–664.

Zhang, Guoqiang, B Eddy Patuwo, and Michael Y Hu (1998). "Forecasting with artificial neural networks:: The state of the art". In: *International journal of forecasting* 14.1, pp. 35–62.