QUANTITATIVE METHODS FOR BIG DATA: **NEURAL NETWORKS**

Universitat Autònoma de Barcelona

Marta Oliva Riera

INTRODUCTION

The aim of this thesis is to analyse possible ways in which big data techniques could be implemented in the field of economics.

OBJECTIVES

- Learn about big data and which of its tools can be used in economics
- Learn how neural networks work
- Set up and train a neural network to predict the results of an economic model

BIG DATA

Big data refers to extremely large datasets that are too complex to be analysed using traditional methods - they require new tools. These new tools can provide economists with more powerful data manipulation, better variable selection and more complex modelling of non-linear relationships.

Machine learning is the field concerned with using algorithms to make computers learn from the data provided to them, in order to improve their performance in a specific task. It has been used for predictive modelling.
Some machine learning tools that can face the overfitting problem are:

- K-fold cross-validation
- Regularization
- Bootstrap, Bagging, Boosting

NEURAL NETWORKS

An **artificial neural network** is a computing system made up of interconnected processing units or neurons. Through parallel processing of the information, it can solve classification and regression problems without the need to prespecify the relationship between variables.
A neuron's basic structure is illustrated in figure 1.

Fig. 1 Neuron k

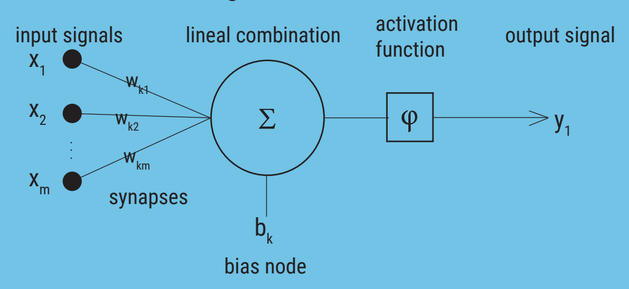
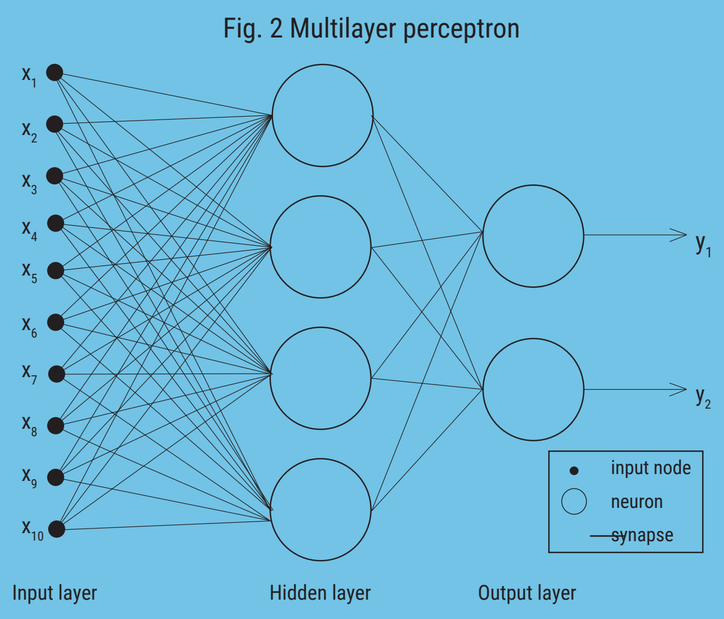


Fig. 2 Multilayer perceptron



The most common type of network architecture is the multilayer perceptron: a network which has hidden layers. The presence of at least one hidden layer allows the network to model more complex relationships.

During the **learning process**, the synaptic weights of the network are adjusted based on the information fed to it. In supervised learning, the network has no knowledge of the environment so the user has to provide it with the "train" subset of data, which represents the desired response.
The adjustment is done according to a learning algorithm and it is repeated until the performance of the network is improved. In error correcting learning, the performance is improved when the difference between the desired and actual output (the error) is minimized.

EMPIRICAL EXERCISE

The **neuralnet** R package was used to train a multilayer feedforward NN, using a data set on wages and other characteristics of individuals. The data set includes 4165 observations across 6 years, and its source is the Ecdat package.
Having normalised the data, it was split into train and test sets which were used to perform the learning process. Two architectures of the neural network were trained: the first with 6 hidden neurons (NN6), the second with 8 and 3 hidden neurons in two layers (NN83).

Fig. 3 Trained NN6

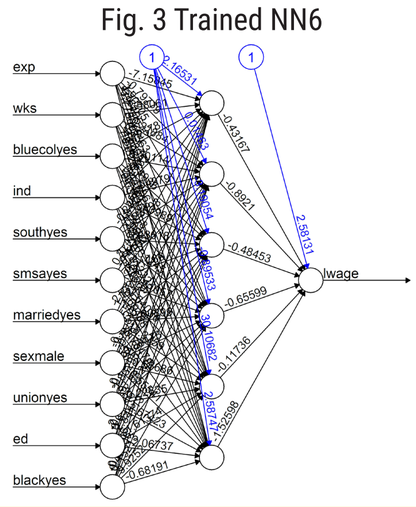
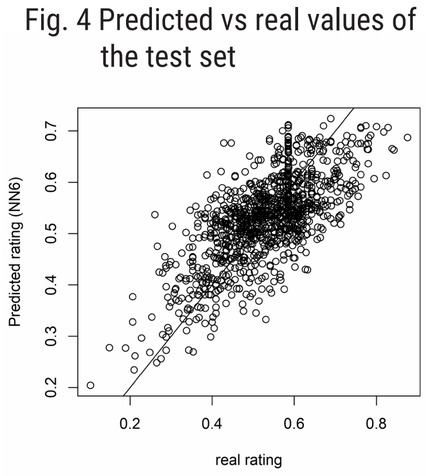


Fig. 4 Predicted vs real values of the test set



The first trained network is represented in figure 3, while figure 4 compares the predictions made by the network to the real values of the test set. As can be seen on the graph, the prediction was adequate. The other specification gave very similar results.
Computing the mean squared error of both specifications, the results were 0.0965 and 0.0925 respectively - there was a little improvement derived from the second model specification.

CONCLUSION

- Big data could become useful to economists if they were to adopt some of the techniques previously mentioned. The dimension and granularity of data present new research opportunities but require more manipulation and summarisation as well as the use of new tools.
  There is also potential for collaboration between economics and machine learning in causal inference. While it has only focused on predictive models, machine learning could provide causal conclusions by predicting the performance of a hypothetical control group.
- Neural network are a method that can be used to model complex relations in large data sets. They are complex but still approachable to someone without a programming background.
  The empirical exercise was very useful to put into practice the concepts and get acquainted with new techniques. However, there is scope for further application of these tools and it would be interesting to use them with larger economic data sets to see what predictions they provide. It could also be interesting to compare its performance to that of an equivalent linear model.

SELECTED BIBLIOGRAPHY

Varian, H., 2014a. Big Data: New Tricks for Econometrics. Journal of Economic Perspectives, 28(2): pp.3-28
Haykin, S., 2009. Neural Networks and Learning Machines. 3rd ed. Pearson Prentice Hall