applied in financial modelling. In particular, attention is paid to value-at-risk estimation by using probabilistic fuzzy systems. A sequential approach is proposed for determining the model parameters, where the location of the antecedent membership functions is determined by using fuzzy clustering while maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The validity of the VaR models obtained is evaluated by using a statistical back-testing method (Kupiec test) based on failure rates.

#49: **Parameter-free feature selection with mutual information**

Presenter: . . . . . Michel Verleysen@Universite Catholique de Louvain, Belgium
Co-authors: . . . . . Damien Francois

Machine learning of high-dimensional data faces the curse of dimensionality, a set of phenomena that limit the performance of the tools. Many limitations come directly from the representation of the data, and not from the analysis tool. It is therefore needed to reduce the data dimensionality. There are basically two ways to do this: either to select features among the original variables, or to project the latter on new ones. Although more general and thus more powerful in theory, projecting features induces a loss of interpretability. On the contrary, by selecting original features, one can come back to the application and interpret which are the relevant factors for the analysis; this is important advantage in many applications. This paper shows how to use Mutual Information (MI) for feature selection. In practice, the MI criterion has to be estimated and the search for possible feature subsets restricted for computation time reasons. It is shown how to use resampling and permutation tests to select optimal parameters for the estimator, and to stop the search procedure in a sound way. It is also shown how to design an estimator of feature subset relevance inspired from the mutual information criterion, with the supplementary advantage to restrict the estimation to a two-dimensional problem.

#87: **Fuzzy text mining and digital obesity**

Presenter: . . . . . Trevor Martin@University of Bristol, UK
Co-authors: . . . . . Yun Shen

The phrase *digital obesity* summarises a range of problems arising from our propensity to generate and retain a rapidly growing volume of data, at web-scale as well as at corporate and personal scales. Much of this data is in text form, but is effectively wasted unless we can find and use the *right* data when needed. Statistical methods help to a degree, but tend to *average out* useful information, as well as suffering from a mismatch between the precisely defined terms used by formal models and the far more subtle and expressive terms used in human communication. Humans communicate using language where the majority of concepts are fuzzy, defined by common usage rather than by necessary and sufficient conditions. The success of fuzzy control is one example where fuzzy set theory enabled computers to work with ill-defined terms such as *hot* and *slow* rather than precise values. Fuzziness enables computers to work with ill-defined concepts, leading to more effective use of text-based information in business and other situations. Although the input information is rarely complete (and may be incorrect) the approximately correct solutions are generally sufficient as well as being easier to compute and understand.

#96: **The estimation of prediction error for neural networks: a simulation study.**

Presenter: . . . . . Simone Borra@University of Rome "Tor Vergata", Italy
Co-authors: . . . . . Agostino Di Ciaccio

One fundamental problem in statistics is that of obtaining an accurate estimate of the prediction error, i.e. the expected loss on future observations, of a learning algorithm trained on the available sample data. This problem has particular relevance every time a very large sample is not available, the underlying distribution is not known and you need to evaluate the prediction error of a non-parametric model which could overfit data. The simplest estimator of prediction error is the Apparent Error defined as the average of the loss function on the training data-set. Apparent error usually produces an optimistic estimate of prediction error because it uses the same data both for training and for evaluation of the model. Using powerful non-linear models, as Neural Networks, it is possible to obtain very small values of Apparent Error, just including more parameters in the model. A way to evaluate the prediction error of the model is to estimate the Optimism, defined as the expected difference between the prediction error and the Apparent Error on new training data, adding it to the Apparent Error. We considered several approaches to prediction error estimation for Neural Networks. In particular, estimators based on Cross-Validation (as Leave-one-out, K-fold cross-validation) and Repeated Cross-Validation (obtained averaging a set of cross-validation estimates on different random split); estimators based on non-parametric Bootstrap (as 0.632 bootstrap and the modified version 0.632+ to take into account situations of severe overfit) and parametric Bootstrap (where the Optimism is proportional to a covariance term estimated by Bootstrap). Using an extensive simulation approach we were able to compare the estimators with respect to different characteristics of data. We considered a regression problem with 1000 data generating distributions showing different level of non-linearity and signal/noise ratio. In each population, we drew 30 samples on which we trained two different NN, calculating also all estimators of prediction error. We generated also a very large sample from each population, to obtain a reliable estimation of the true prediction error for each NN. Finally, we compared all prediction error estimators on the bases of bias and variability. We obtained some interesting suggestions about the efficiency of the different prediction error estimators with respect to the s/n ratio and the neural network complexity.

#114: **Application of neural networks and support vector machines to pricing European options**

Presenter: . . . . . Chris Charalambous@University of Cyprus, Cyprus

Artificial Neural networks (ANN), as discipline, studies the information processing capabilities of networks made up of simple processors which are in some way connected with different strengths (weights) like the living neurons of the brain. During the