

Coventry University Repository for the Virtual Environment
(CURVE)

Author names: Jayne, C. , Lanitis, A. and Christodoulou, C.

Title: Neural network methods for one-to-many multi-valued mapping problems.

Article & version: Post-print

Original citation & hyperlink:

Jayne, C. , Lanitis, A. and Christodoulou, C. (2011) Neural network methods for one-to-many multi-valued mapping problems. *Neural Computing & Applications*, volume 20 (6): 775-785.

<http://dx.doi.org/10.1007/s00521-010-0483-4>

Publisher statement:

The final publication is available at www.springerlink.com.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Available in the CURVE Research Collection: July 2012

<http://curve.coventry.ac.uk/open>

One-To-Many Multi-Valued Mapping Problems: A Comparative Neural Network Evaluation

Chrisina Draganova¹, Andreas Lanitis², Chris Christodoulou³

¹ *University of East London, 4-6 University Way, London E16 2RD, UK*

² *Department of Multimedia and Graphic Arts, Cyprus University of Technology, 31 Archbishop Kyprianos Street, P. O. Box 50329, 3603 Lemesos, Cyprus*

³ *Department of Computer Science, University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, 1678 Nicosia, Cyprus*

{c.draganova@uel.ac.uk, andreas.lanitis@cut.ac.cy, cchrist@cs.ucy.ac.cy}

Abstract.

In this paper we investigate the applicability of neural-network based methods in predicting the values of multiple parameters, given the value of a single parameter within a particular problem domain. In this context the input parameter may be an important source of variation that is related to the remaining sources of variation within a multivariate distribution, with a complex mapping function. The definition of the relationship between the variables of a multivariate distribution and a single source of variation allows the estimation of the values of multiple variables given the value of the single variable, addressing in that way an ill-conditioned one-to-many mapping problem. As part of our investigation two problem domains are considered: predicting the values of individual stock shares, given the value of the general index and predicting the grades received by high school pupils, given the grade for a single course. With our work we compare the performance of standard neural network based methods and in particular Multilayer Perceptrons (MLPs) and Radial Basis Functions (RBFs) as well as Mixture Density Networks (MDNs) and a latent variable method, the General Topographic Mapping (GTM). According to the results, MLPs and RBFs outperform MDNs and the GTM for these one-to-many mapping problems.

Keywords: *Stock Price Prediction, Exam marks prediction, Neural Networks, Multivariate Statistics, One-to-Many Mapping.*

1 Introduction

One-to-many mapping problems involve the estimation of the values of multiple parameters given the value of a single parameter within a specific problem domain. In general, one-to-many mapping problems are ill-conditioned, requiring the use of dedicated techniques that utilise prior knowledge related to the relationship between the input and output variables in an attempt to formulate an optimized mapping function that allows the estimation of the values of the output parameters with a reasonable accuracy. With our work we aim to investigate the use of different methods for defining a mapping associating a specific source of variation within a distribution and a given representation of this data distribution. As part of our performance evaluation framework we assess the performance of a number of one-to-many mapping methods in two case studies.

The first case study is related to the definition of the relationship between the index value of twenty stocks included in the FTSE 100 UK (www.ftse.com/Indices/UK_Indices/index.jsp) and the daily individual stock prices over a three year time period. We implement and test methods that learn the relationship between the daily general index value and the corresponding individual daily stock prices of twenty of the FTSE 100 UK stocks with largest volume that have available data for at least three consecutive years. Once the mapping is learned we attempt to predict the daily stock prices of each share given the value of the general index. This application can be very useful for efficient portfolio management. There are numerous studies [1-12] investigating problems related to predicting stock prices based on neural networks methods due to the ability of neural networks to model nonlinear systems. However, the problem of finding the relationship between the general index and the prices of set of shares in the context of on-to-many mapping has to the best of our knowledge never been considered.

The second case study is related to predicting the grades received by pupils in a number of different subjects, based on the grade received in a single subject or an average grade. For the needs of the experiment we used the results of examinations taken by Cypriot high school graduates as part of a University admission process (all results are publicly available at <http://www.moec.gov.cy/ypexams/apotelesmata.html>). Each year about 10000 candidates take these examinations but only a subset of all candidates is examined on the same set of four subjects according to the discipline in which candidates aim to pursue a University Degree. For the needs of our experiments a popular combination of the subjects is selected: Modern Greek, Biology, Mathematics and Physics. The grades received by candidates are expressed in a 1 (lowest) to 20 (highest) grade. As part of our experimental evaluation, we run one to many mapping experiments that aim to either use the average grade obtained by a candidate as the basis for predicting the grades for the remaining four courses or use the grade of a single subject as the basis for predicting the grade for the remaining three courses. The ability to predict the grades received by a pupil given his/her grade in a single subject can be used for assessing how the analytic knowledge and skills gained by a pupil in a single course can play a key role in his/her performance in other courses. Research in predicting student's grades or academic performance based on neural networks has been done in a number of studies (e.g. [13-15]) but not in the form of mapping a single subject grade to a set of several other grades.

As part of our experimental evaluation process we investigate the following neural network-based methods: Multilayer Perceptron (MLP) [16], Radial Basis Functions [17], Mixture Density Networks (MDN) [18, 19] and the non-linear latent variable method Generative Topographic Mapping (GTM) [20]. The performance of the methods mentioned above is compared in the problems of stock share price prediction and student grade prediction, so that the applicability of each of the methods in one-to-many mapping problems is quantified.

The rest of the paper is organised as follows: in section 2 we present an overview of the relevant literature; in section 3 we describe the case studies under investigation, the experiments performed and the visual and quantitative results obtained and in section 4 we present our conclusions.

2 Literature Review

The problems of stock price prediction [1-12] and grade prediction [13-15] received considerable attention in the literature. However, in the case of stock-price predictions, previous studies treat the problem as a typical many-to-one problem rather than as an one-to-many problem. Similarly previous studies on student grade prediction [13-15] has been used for predicting the overall performance of students rather than predicting the grade received by students in a number of courses.

Saad et al. [21] performed a comparative evaluation of Time Delay, Probabilistic and Recurrent neural networks (TDNN, PNN and RNN respectively) in predicting stock market daily closing prices using time series data of the stock in question as input. As they showed, all different NNs are equally feasible and they suggested that the primary preference should be the most convenient network. Yudong and Lenan [2] integrated an improved bacterial chemotaxis optimisation into an MLP Artificial Neural Network (ANN) with the backpropagation learning algorithm for predicting the S&P 500 stock market index using time series data for the said index reporting good performance in terms of prediction accuracy and training time. Apart from the above a huge number of other researchers used time series data with ANNs for forecasting either the general index value of stock markets or of a value of a particular stock (see for example [3-12]); others used other factors like trading volume [22], while others attempted to forecast the volatility of the stock price index [23]. All of the above models which are either univariate (using previous time series data for constructing a predictor) or multivariate (using apart from past time series data, additional information for constructing a forecaster) are approaching the stock price prediction problem from a typical many-to-one point of view, which is diametrically different to our approach in this paper where we are treating the problem as an one-to-many problem.

Hardgrave et al [13] and Gorr et al [15] use neural-network models for predicting the grade point average (GPA) obtained by graduate students, given the performance of students in several courses during their undergraduate studies and also given demographic data related to each student. The problems considered by Hardgrave et al [13] and Gorr et al [15] are a typical many-to-one mapping scheme – a problem that is widely investigated by numerous researchers. Hardgrave et al [13] and Gorr et al [15] conclude that neural-network based approaches have the potential to perform better than traditional statistical modeling techniques in the task of grade prediction. Along similar lines Sheel et al [24] train a neural network with inputs the high school GPA, the grade received at mathematics SAT and the final grade in an advanced algebra course. The output of the trained neural network predicts the optimum mathematics placement level. Sheel et al [24] prove that neural network approaches are more accurate in predicting the optimum mathematics placement class when compared with classical statistical approaches, such a discriminant analysis. More recently Paliwal and Kumar [14] investigate the use of neural network and regression techniques for predicting the performance of graduate Business School students, in an attempt to establish an accurate student admission procedure. In this context the aim is to classify students as successful or not based on a number of input parameters that include the undergraduate grade point average, test score of an admission test, work experience, age, gender, and metrics derived from the reference letters, group discussion and personal interview. According to the

results neural network methods outperform standard regression analysis methods. In contrast to the approaches reported in [13-15, 24] that involve a classical many-to-one mapping problem, our work in the area of grade prediction takes the form of one-to-many mapping scheme where we aim to estimate the grades received by students in multiple subjects given the grade of a single subject.

There exist well-established neural network methods for solving the mapping approximation problem such as the Multilayer Perceptron (MLP) [16] and Radial Basis Functions (RBF) [17]. The aim of the training in these methods is to minimize a sum-of-square error function so that the outputs produced by the trained networks approximate the average of the target data, conditioned on the input vector [18]. It is reported in [19] and [20], that these conditional averages may not provide complete description of the target variables especially for problems in which the mapping to be learned is multi-valued and the aim is to model the conditional probability distributions of the target variables [19]. In our case despite the fact that we have a multi-valued mapping we aim to model the conditional averages of the target data, conditioned on the input that represents a source of variation within this distribution. The idea is that when we change the value of the input parameter, the mapping that is defined will give typical representation of the target parameters exhibiting the isolated source of variation.

Bishop [18, 19] introduces a new class of neural network models called Mixture Density Networks (MDN), which combine a conventional neural network with a mixture density model. The mixture density networks can represent in theory an arbitrary conditional probability distribution, which provides a complete description of target data conditioned on the input vector and may be used to predict the outputs corresponding to new input vectors. Practical applications of feed forward MLP and MDN to the acoustic-to-articulatory mapping inversion problem are considered in [25]. In this paper, it is reported that the performance of the feed-forward MLP is comparable with results of other inversion methods, but that it is limited to modelling points approximating a unimodal Gaussian. In addition, according to [25], the MLP does not give an indication of the variance of the distribution of the target points around the conditional average. In the problems considered in [19] and [25], the modality of the distribution of the target data is known in advance and this is used in selecting the number of the mixture components of the MDN.

Other methods that deal with the problem of mapping inversion and in particular mapping of a space with a smaller dimension to a target space with a higher dimension are based on latent variable models [26]. Latent variables refer to variables that are not directly observed or measured but can be inferred using a mathematical model and the available data from observations. Latent variables are also known as hidden variables or model parameters. The goal of a latent variable model is to find a representation for the distribution of the data in the higher dimensional data space in terms of a number of latent variables forming a smaller dimensional latent variable space. An example of a latent variable model is the well-known factor analysis, which is based on a linear transformation between the latent space and the data space [18]. The Generative Topographic Mapping (GTM) [20] is a non-linear latent variable method using a feed-forward neural network for the mapping of the points in the latent space into the corresponding points in the data space and the parameters of the model are determined using the Expectation-Maximization (EM) algorithm [27]. The practical implementation of the GTM has two potential problems: the dimension of the latent space has to be

fixed in advance and the computational cost grows exponentially with the dimension of the latent space [28].

Density networks [29] are probabilistic models similar to the GTM. The relationship between the latent inputs and the observable data is implemented using a multilayer perceptron and trained by Monte Carlo methods. The density networks have been applied to the problem of modelling a protein family [29]. The biggest disadvantage of the density networks is the use of the computer-intensive sampling Monte Carlo methods, which do not scale well when the dimensionality is increased.

Even though the problems we consider in this paper bear similarities with the problem of sensitivity analysis with respect to neural networks, there are also distinct differences. In sensitivity analysis the significance of a single input feature to the output of a trained neural network is studied by applying that input, while keeping the rest of the inputs fixed and observing how sensitive the output is to that input feature (see for example [30] and references therein). In the problems investigated in this paper, we do not have a trained neural network, but the index based on the values of 20 stocks in the first problem and in the second problem the average grade based on four subjects or the grade of one subject in relation to three other subjects. Based on our knowledge of the application, i.e., the index or one single subject/average grade, we isolate a specific source of variation and carry out an one-to-many mapping between that isolated source and the model (which is a multivariate data distribution). More specifically, the model refers to all the 20 stock values in the first case study and all four subject grades in the second case study. This allows us to analyse the variation of the isolated source within the model.

3 Experiments and Results

3.1 Experiments

For the experiments related to stock price prediction described in this paper we have used the historical daily prices, expressed in pence, available at uk.finance.yahoo.com. Twenty stocks have been selected from those that have the largest volume and that have their daily prices between 16/12/2003 and 10/12/2007. Precisely the set of selected stocks includes: BA, BARC, BLT, BP, BT, CW, FP, HBOS, HSBA, ITV, KGF, LGEN, LLOY, MRW, OML, PRU, RBS, RSA, TSCO and VOD. The daily index values for these 20 stocks have been calculated using the method described in [31] with a starting index point set to 1000. Among the 1019 available samples, 719 randomly selected samples are used for training and the remaining 300 samples are used for testing the performance of the methods.

For the second case study related to the prediction of grades we run five different experiments predicting the grades of three of the selected subjects given the grade of a fourth subject or predicting the grades of the four subjects given the average grade between them. For example, we find the mapping between the grade of Modern Greek (subject code 1) and the grades of Biology (subject code 21), Mathematics (subject code 37) and Physics (subject code 38). For our experiments we use 433 samples each one corresponding to the grades received by a candidate who was examined in the four subjects. The samples were split to

training set of 333 randomly selected samples and a testing set of 100 randomly selected. In order to produce experimental results for a large number of test cases, the process of formulating training and test sets was repeated four times using different randomly selected train and test sets and the results presented here are averaged over the four trials.

We first train neural network models using the MLP with the scaled conjugate gradient algorithm [32], the RBF and MDN methods. In the case of the stock price prediction problem, the inputs for the neural network model are the numerical values of the daily general index value and the output corresponds to the prices of the 20 stocks for the FTSE case study. In the case of the grade prediction problem the inputs for the neural network model are the values of one subject exam mark and the outputs are the values of the remaining subject exam marks. As part of the grade prediction problem, the mapping between the average of the four subject marks and the individual marks is also considered. In the MLP model, the network has one input node, one hidden layer with hyperbolic tangent (tanh) activation function and an output layer with linear activation function, since the problem we consider is a regression problem [33]. In the case of RBF similarly to the MLP, the input layer has one node, the output layer has linear outputs and the hidden layer consists of nodes (centres) with Gaussian basis functions. The Gaussian basis function centres and their widths are optimised by treating the basis functions as a mixture model and using the Expectation-Maximisation (EM) algorithm for finding these parameters. The number of hidden nodes in the MLP and RBF networks and the learning rate in the MLP network are set empirically. We also set empirically the number of hidden nodes and kernel functions (mixture components) in the MDN model. Theoretically by choosing a mixture model with a sufficient number of kernel functions and a neural network with a sufficient number of hidden units, the MDN can approximate as closely as desired any conditional density. In the case of discrete multi-valued mappings the number of kernel functions should be at least equal to the maximum number of branches of the mapping. We performed experiments using up to five kernel functions.

We have also carried out an experiment for isolating one latent variable using the GTM. The GTM models consist of an RBF non-linear mapping of the latent space density to a mixture of Gaussians in the data space of parameters (20 stock prices or a set of 3, 4 subjects). The models are trained using the EM algorithm. After training we use the RBF mapping to obtain the parameters corresponding to several values of the latent variable. We show the variation of the latent variable reflected on the stock prices (stock problem) or on the individual grades for each subject (grades problem).

3.2 Results

In the following sections we present the results of the comparative evaluation procedure for the two case studies considered.

3.2.1 Stock prices problem

Table 1 represents the quantitative results for each method used, expressed as the mean error between actual and predicted stock prices in pence over the considered period of time. The last row of Table 1 gives the overall average percentage error between actual and predicted stock prices of all 20 shares. Figure 1 illustrates the graphical results for the actual and the model output prices of one of the stocks - ITV obtained with the MLP, RBF, MDN and GTM methods. These graphical

results show the variation of the index value against the predicted and actual prices of the ITV stock. The graphs indicate clearly that a single value of the input variable (stock index) may correspond to a range of values of the stock price. It is important that the methods used learn the most probable stock price index, among the candidate values. The visual results in figure 1 show that the MLP and RBF methods approximate in a reasonable way the variation in a single stock price given the value of a stock index.

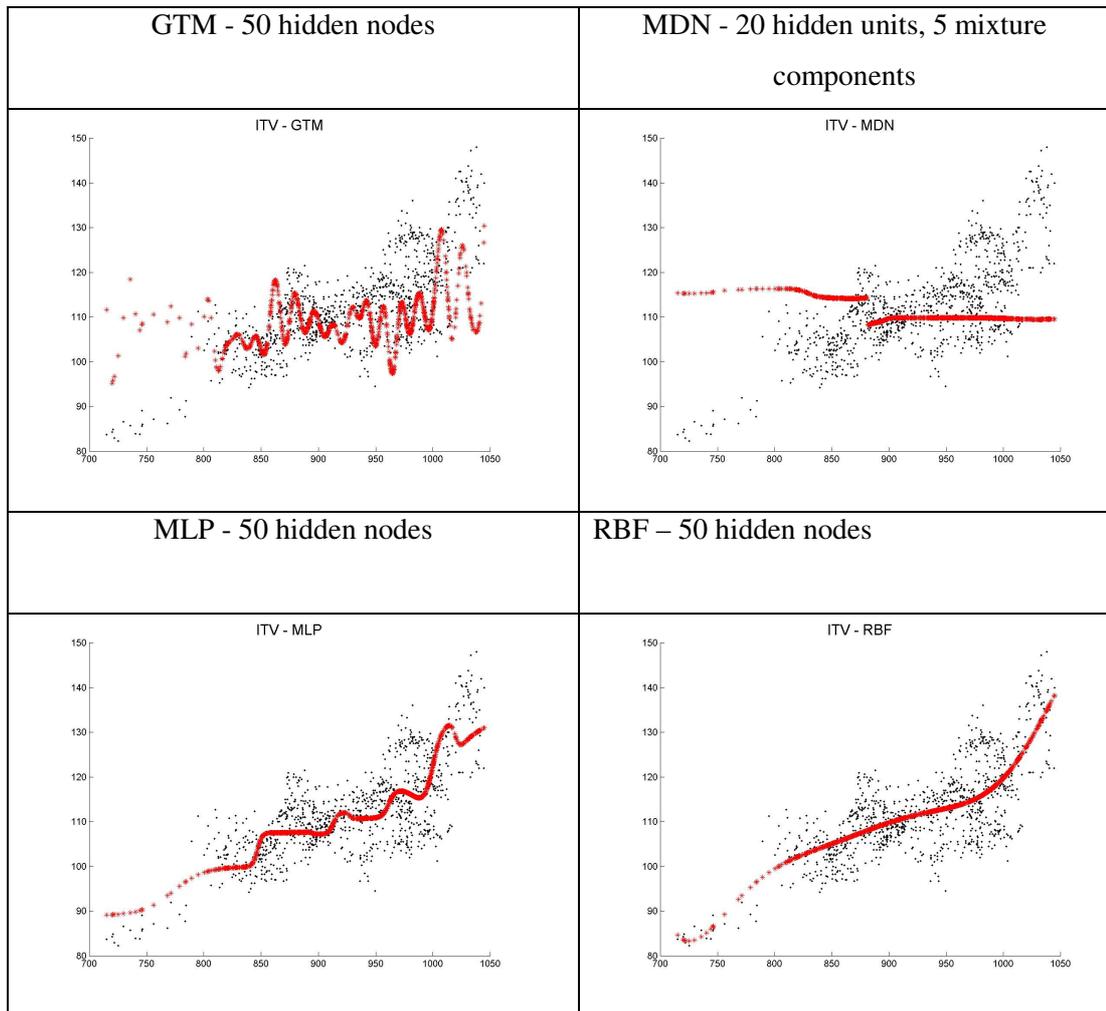
The quantitative results corresponding to the MLP and RBF models are comparable. The results obtained with the MDN method did not produce better representation of the data which again can be explained with the large dimensionality of the problem. In the case of the GTM method, although the quantitative results are worse than those obtained with the other methods, the graphical results show that the general trend of the actual prices is captured, demonstrating therefore the potential of the GTM method for modeling the distribution of the stock prices in terms of one latent variable. In general the differences between the performance on the train and test sets are minimal, suggesting that the methods can generalize well on previously unseen data.

Table 1. Mean error between actual and predicted prices in pence of the 20 listed shares with different methods

Share	Method							
	MLP		RBF		MDN		GTM	
	train	test	train	test	train	test	train	test
BA	45.59	48.31	49.03	51.98	56.47	55.09	100.38	100.61
BARC	41.67	40.82	44.05	43.03	51.80	52.37	78.73	84.28
BLT	141.99	155.04	156.37	163.59	185.98	199.87	279.75	269.31
BP	40.68	44.96	45.60	48.03	52.42	49.80	73.09	72.03
BT	19.60	19.84	24.34	23.90	26.98	25.89	40.11	41.44
CW	14.85	14.65	17.88	17.54	21.99	21.74	26.05	23.63
FP	14.58	14.89	15.00	1.94	19.57	19.93	22.58	22.92
HBOS	68.62	68.80	70.60	68.32	89.24	92.36	132.86	138.61
HSBA	31.19	31.92	31.15	33.41	37.04	36.69	59.35	56.19
ITV	4.65	4.90	5.07	5.02	7.41	8.01	10.46	9.54
KGF	18.42	19.17	19.28	18.36	26.24	29.41	39.88	37.69
LGEN	9.76	10.10	10.94	10.47	12.15	11.42	19.48	21.53
LLOY	28.25	28.57	30.64	29.23	32.15	31.50	52.72	57.09
MRW	23.51	23.77	28.35	28.58	31.33	30.41	48.61	52.07
OML	14.95	16.53	15.48	15.99	19.13	18.83	33.61	34.98
PRU	49.82	53.16	57.75	53.61	61.57	58.89	95.44	104.90
RBS	26.58	25.89	28.72	25.63	35.56	37.19	52.56	55.49
RSA	13.12	13.67	13.87	13.51	19.57	15.67	38.26	41.36
TSCO	28.57	28.86	32.27	34.31	37.74	36.87	54.11	53.47
VOD	8.07	8.61	8.37	8.49	10.94	12.03	24.35	24.77
Average error	32.22	33.62	35.24	35.40	41.58	42.20	64.12	65.10
Average % error	8.8	9.0	9.5	9.9	10.9	10.9	17.9	17.9

The mean error = $(\sum_{i=1,n} \text{abs}(y_i - a_i))/n$, where y_i is the predicted and a_i is the actual price of the shares, n is the total number days over which the share prices are predicted.

Fig. 1. Sample graphical result for the variation of the index value reflected on the ITV stock price; the solid lines and the scattered dots indicate the predicted and actual stock prices respectively corresponding to the index values



3.2.2 Exam grades problem

Tables 2, 3, 4, 5 and 6 represent the quantitative results for each method used, expressed as the mean error between actual and predicted subject mark over the train and test sets and the average percentage error. The Figures 2 and 3 illustrate the graphical results for the actual and the model output subject marks obtained with the MLP, RBF, MDN and GTM methods.

Similar to the stock prices case study the graphical and quantitative results corresponding to the MLP and RBF models are comparable and the GTM method gives worse quantitative results but captures the general trend of the actual grades. The results obtained with the MDN method give also comparable representation of the data since in this case the problem is with smaller dimension. It is not a surprise that according to the results the most accurate prediction of individual grades is done when the average grade is used as an input. When the grades of a single subject are used, the most accurate predictions were observed in the cases that the grades for Biology (subject code 21) and Physics (subject code 38), were used as the input.

Table 2. Predicting Subjects 1, 21, 37 given Subject 38 as input

Subject Code	Method							
	MLP		RBF		MDN		GTM	
	Train	test	train	test	train	Test	train	test
1	1.47	1.54	1.49	1.58	1.58	1.60	1.91	2.01
37	1.94	2.13	1.98	2.10	2.15	2.15	2.69	2.81
38	1.97	2.08	1.99	2.06	2.07	2.11	2.73	2.82
Average Error	1.79	1.92	1.82	1.91	1.93	1.95	2.44	2.55
Average % error	9.0	9.6	9.1	9.6	9.7	9.8	12.2	12.8

Table 3. Predicting Subjects 1, 21, 38 given Subject 37 as input

Subject Code	Method							
	MLP		RBF		MDN		GTM	
	Train	test	train	test	train	Test	train	test
1	1.45	1.49	1.45	1.51	1.47	1.56	2.00	1.98
37	2.25	2.47	2.33	2.41	2.42	2.58	3.40	3.36
38	1.89	1.99	1.94	1.97	2.03	2.06	3.19	3.21
Average Error	1.86	1.98	1.91	1.96	1.98	2.07	2.86	2.85
Average % error	9.3	9.9	9.6	9.8	9.9	10.4	14.3	14.3

Table 4. Predicting Subjects 1, 37, 38 given Subject 21 as input

Subject Code	Method							
	MLP		RBF		MDN		GTM	
	Train	test	train	test	train	Test	train	test
1	1.42	1.48	1.42	1.48	1.52	1.54	1.85	1.89
37	2.13	2.30	2.13	2.27	2.42	2.57	3.07	3.14
38	1.79	1.89	1.79	1.92	1.98	2.06	2.91	2.96
Average Error	1.78	1.89	1.78	1.89	1.97	2.06	2.61	2.66
Average % error	8.9	9.5	8.9	9.5	9.9	10.3	13.1	13.3

Table 5. Predicting Subjects 21, 37, 38 given Subject 1 as input

Subject Code	Method							
	MLP		RBF		MDN		GTM	
	Train	test	train	test	train	Test	train	test
21	2.64	2.85	2.70	2.83	2.93	3.11	1.85	3.49
37	2.62	2.79	2.66	2.78	2.89	3.05	3.07	3.40
38	2.54	2.67	2.55	2.67	2.79	2.83	2.91	3.33
Mean Error	2.60	2.77	2.64	2.76	2.87	3.00	2.61	3.41
Average % error	13.0	13.9	13.2	13.8	14.4	15.0	13.1	17.1

Table 6. Predicting Subjects 1, 21, 37, 38 given the average as input

Subject Code	Method							
	MLP		RBF		MDN		GTM	
	Train	test	train	test	train	Test	train	test
1	1.17	1.25	1.17	1.28	1.20	1.28	1.83	1.77
21	1.25	1.51	1.26	1.59	1.36	1.70	2.47	2.46
37	1.32	1.41	1.34	1.52	1.45	1.57	2.57	2.76
38	1.22	1.27	1.23	1.34	1.28	1.33	2.45	2.62
Average Error	1.24	1.36	1.25	1.43	1.32	1.47	2.33	2.40
Average % error	6.2	6.8	6.3	7.2	6.6	7.4	11.7	12.0

Fig. 2. Sample Graphical Results: Predicting Subjects 1, 21, 37 on the basis of Subject 38 as input. The figures show the results for Subject 1

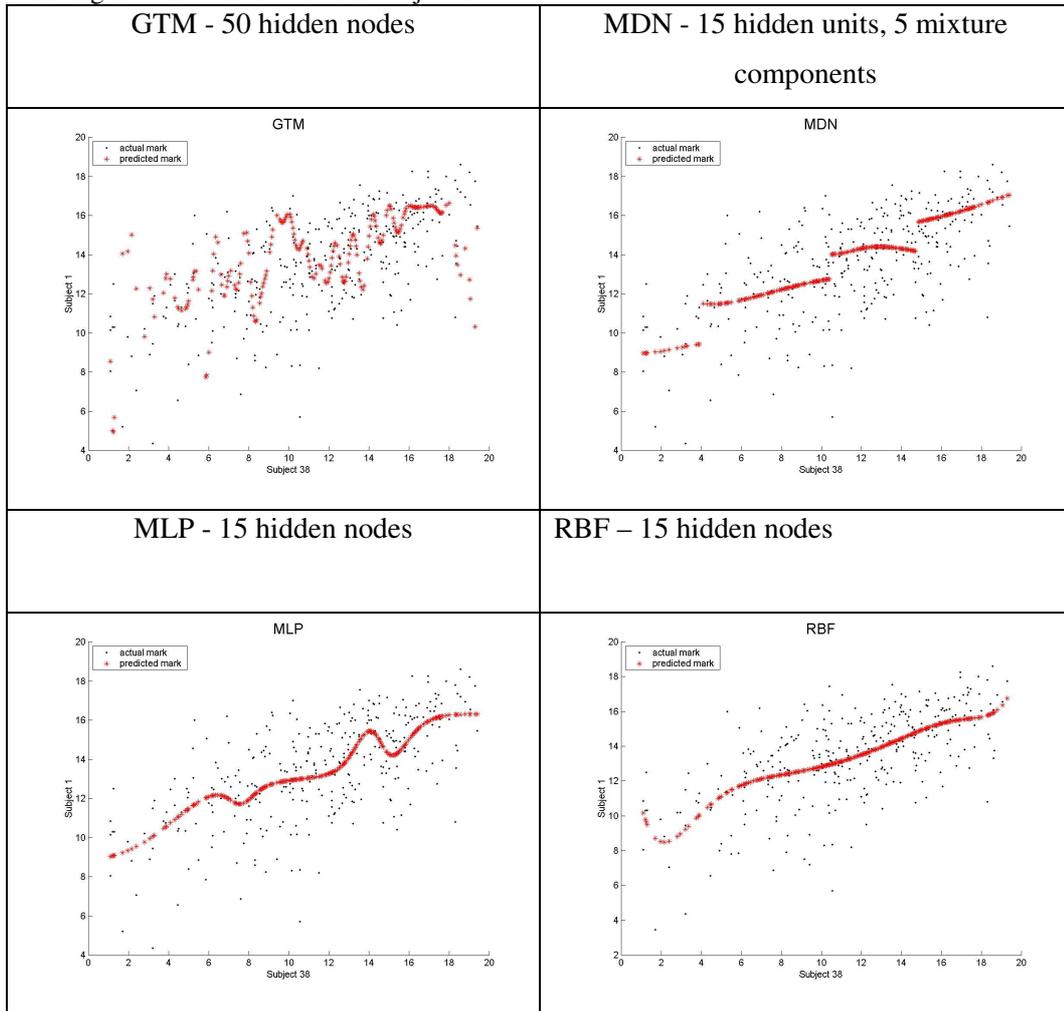
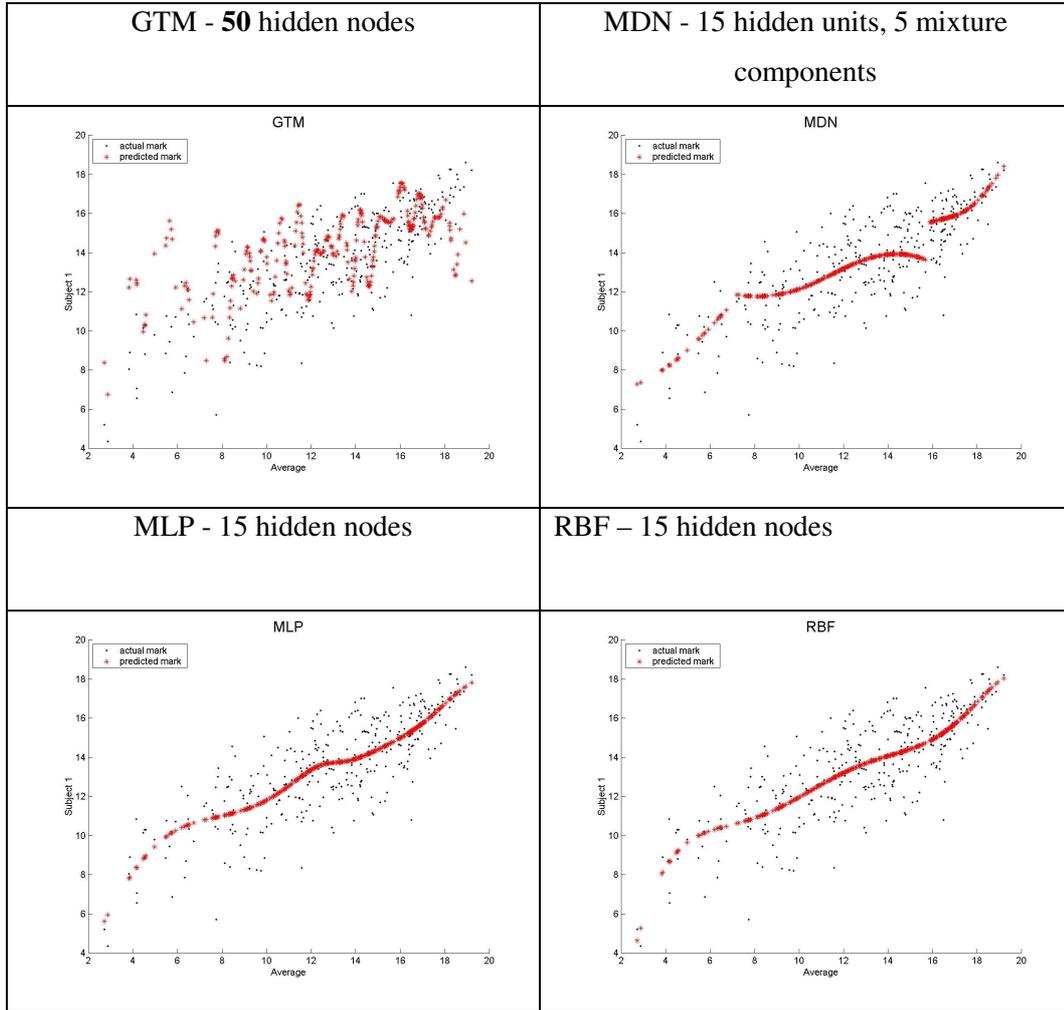


Fig. 3. Sample Graphical Results: Predicting Subjects 1, 21, 37, 38 on the basis of Average grade as input. The figures show the results for Subject 1



4 Conclusions

The performance of different neural network paradigms in ill-conditioned one-to-many multi-valued mapping problems is evaluated in this paper. The 'one' is a source of variation representing a quantity of interest within a multivariate data distribution of a problem domain and the 'many' represents the data distribution itself. In particular, we consider two one-to-many problems. The first problem relates to the definition of a mapping function relating the general index value among twenty shares, to the prices of the twenty individual shares themselves. The second problem relates to the mapping of one grade or an average grade received by a high school graduate to the grades of three or four subjects respectively. The mapping function we aim for gives a typical representation of the data distribution exhibiting the variation of the specific quantity. Given that the target space in such a mapping has a higher dimension than the input space, it follows that the target output value from one input value is not unique. The neural networks we evaluated for these one-to-many problems include MLPs, RBFs, MDNs and GTMs.

The results of our experiments demonstrate the potential of using neural networks trained with the MLP and RBF methods for isolating sources of

variation and generating typical representations of the corresponding data distributions in the considered case studies. With the neural network approach we do not make any assumptions about the mapping function; the neural networks are learning the complex mapping between the desired attributes and the parameters related the specific applications. In both test cases the quantitative results obtained with the MLP and RBF are similar but the best result is achieved with the MLP method. The graphical results obtained with these methods are also similar. The MLP and RBF methods give the conditional averages of the target data conditioned on the input vectors and as expected they do not give a complete description of the target data reported in [19, 25]. For the problems we are addressing it is sufficient to define a mapping that generates typical samples of the data distribution (and not its entire variance) given specific values of the desired source of variation. This makes our results with the MLP and RBF (which are relatively simple methods, compared to MDN and GTM) not only acceptable but quite good for the type of inversion problems we are addressing, compared to the MLP results for the acoustic-to-articulatory inversion mapping reported in [25, 34]. For this one-to-many problem considered and also the problem of reconstructing the same spectrum from different spectral line parameters [11], the entire variance of the distribution is required. It has to be noted also that for the one-to-many problems considered in our paper the training algorithm for the MLP does not have to be modified as suggested by Brouwer [35], resulting in increased complexity. To the best of our knowledge RBFs have not previously been specifically used for one-to-many problems.

The MDN [18, 19] can give a complete description of the target data conditioned on the input vector provided that the number of mixture components is at least equal to the maximum number of branches of the mapping. The experiments carried out in the considered case studies demonstrate that in problems for which the modality of the distribution of the target data is very large and not known, the application of the MDN leads to a large number of mixture components and outputs. Therefore it does not provide the desired type of mapping, which explains the poor results obtained with the MDN method. In particular, these results do not show the desired variation of the quantity of interest in the stock prices case study, they do not show complete representation of the target data space. In the case of grade prediction where the number of variables are limited when compared to the stock price prediction problem, the performance of MDN is not much worse than the performance of MLP and RBF, indicating that when dealing with a small number of variables MDN can be more applicable.

The GTM [20] is used to map points in the latent variable interval to points in the target data space and our results with this method actually show this mapping of one latent variable to the corresponding data spaces. It has to be noted though, that in order to isolate a specific source of variation, we need to find all latent variables which leads to a problem that is not computationally feasible [28]. In the stock price case study the isolated latent variable might be a different source of variation and not necessarily the desired index value. When dealing with s distribution with fewer and more distinct degrees of freedom (like the grade prediction problem) the performance of GTM is close to the performance of the MLP and RBF, indicating that in the cases that computational complexity is not hindering the application of GTM, there is potential in using GTM and for one-to-many mapping problems.

The framework presented in the paper can be potentially useful in various applications involving multivariate distributions. With regards to the stock prices application, it is well established that trends of the general index can be easily predicted, unlike price fluctuations at share level. Therefore the ability to infer individual stock prices based on the general index value can be an invaluable tool for efficient portfolio management and prediction of the behavior of individual shares. In the case of the grades problem, the techniques described in this paper could be used for predicting the performance of students based on a single course enabling in that way the examination of candidates on a smaller number of courses and the determination of the grades for the remaining courses based on one-to-many mapping schemes. Analysis based on the framework described in this paper can also be useful for assessing the quality of a syllabus and delivery of a course. A well designed and well delivered course will provide to students the necessary analytical knowledge and training that will enable them to succeed in different courses hence the accuracy of the prediction of the grades received by candidates in different courses, based on the grade received on a single subject, can be used for assessing the course itself.

Our future work in this area will concentrate on three distinct directions: (i) evaluation of the usefulness of the results reported in real life-application related to portfolio management, performance evaluation of students and assessment of courses syllabi (ii) investigation of the performance of other methods in one-to-many mapping problems (iii) applications to other problem domains where the proposed framework is applicable.

References

1. Chang P-C, Liu C-H, Lin J-L, Fan C-Y, Ng C S P (2009) A neural network with a case based dynamic window for stock trading prediction. *Expert Syst Appl* 36:3:6889-6898
2. Yudong Z, Lenan W (2009) Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst Appl* 36:5:8849-8854
3. Refenes A N, Zapranis A, Francies G (1994) Stock performance modeling using neural networks: a comparative study with regression models. *Neural Networks* 5:961-970
4. Castiglione F (2001) Forecasting price increments using an artificial neural network. *Advances in Complex Systems* 4:1:45-56
5. Kim S H, Chun S H (1998) Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index. *Int J Forecasting* 14:323-337
6. Kim K-J (2003) Financial time series forecasting using support vector machines. *Neurocomputing* 55:307-319.
7. Sitte R, Sitte J (2000) Analysis of the predictive ability of time delay neural networks applied to the S& P 500 time series. *IEEE T Syst Man CY C* 30:4:568-572
8. Kaastra I, Boyd M (1996) Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10:215-236.
9. Tino P, Schittenkopf C, Dorffner G (2001) Financial volatility trading using recurrent neural networks. *IEEE T Neural Networ* 12:4:865-874
10. de Faria E L, Albuquerque M P, Gonzalez J L, Cavalcante J T P, Albuquerque M P. (2009) Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Syst Appl* 36:12506-12509
11. Chen A-S, Leung M T, Daouk H (2003) Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Comput Oper Res* 30:901-923
12. Pai P-F, Lin C-S (2005) A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*. 33:6:497-505
13. Hardgrave B C, Wilson R L, Walstrom K A (1994) Predicting graduate student success: A comparison of neural networks and traditional techniques. *Comput Oper Res* 21:3:249-263

14. Kumar U A, (2009) A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Syst Appl* 36:4:7865-7872
15. Corr W L, Nagin D, Szczypula J, (1994) Comparative study of artificial neural network and statistical models for predicting student grade point averages. *Int J Forecasting* 10:1:17-34
16. Rumelhart D E, Hinton D E, Williams R J (1986) Learning representations by back-propagation errors. *Nature* 323:533-536
17. Powell M J D (1985) Radial basis functions for multivariable interpolation: A review. IMA Conference on Algorithms for the approximation of Functions and Data, RMCS, Shrivenham, England, pp 143-167
18. Bishop C M (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, New York, Oxford
19. Bishop C M (1994) Mixture Density Networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University, research.microsoft.com/~cmbishop/downloads/Bishop-NCRG-94-004.ps. Accessed 28 January 2010
20. Bishop C M, Svensén M, Williams CKI (1998) GTM: The Generative Topographic Mapping. *Neural Comput* 10:1:215-234
21. Saad E W, Prokhorov D V, Wunsch D C (1998) Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks. *IEEE T Neural Networ* 9:1456-1470
22. Zhu X, Wang H, Xu L, Li H (2008) Predicting stock index increments by neural networks: The role of trading volume under different horizons. *Expert Syst Appl* 34:4: 3043-3054
23. Roh T H (2007) Forecasting the volatility of stock price index. *Expert Syst Appl* 33:4:916-922
24. Sheel S J, Vrooman D, Renner R S, Dawsey S K (2001) A Comparison of Neural Networks and Classical Discriminant Analysis in Predicting Students Mathematics Placement Examination Scores. V.N. Alexandrov et al. (Eds.): ICCS 2001, LNCS 2074, pp 952–957
25. Richmond K (2001) Mixture Density Networks, Human articulatory data and acoustic-to-articulatory inversion of continuous speech. URL www.cstr.ed.ac.uk/downloads/publications/2001/Richmond_2001_a.ps Accessed 28 January 2010
26. Bartholomew D J (1987) *Latent Variable Models and Factor Analysis*. Charles Griffin & Company Ltd, London
27. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B Met* 39:1:1-38
28. Carreira-Perpinan M A (1999) One-to-many mappings, continuity constraints and latent variable models. In: Proc. IEE Colloquium on Applied Statistical Pattern Recognition, Birmingham, pp 14/1-14/6
29. MacKay D J C, Gibbs M N (1997) Density networks. In: Proceedings of Society for General Microbiology. Edinburgh
30. Zeng X, Yeung D S (2003) A Quantified Sensitivity Measure for Multilayer Perceptron to Input Perturbation. *Neural Comput* 15:183-212
31. FTSE Guide to UK Calculation Methods http://www.ftse.com/Indices/UK_Indices/Downloads/uk_calculation.pdf# Accessed 28 January 2010
32. Møler M A (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 6:4:525-533
33. Bishop C M (2006) *Pattern Recognition and Machine Learning*. Springer Verlag, Berlin
34. Carreira-Perpiñán M Á (2001) Continuous latent variable models for dimensionality reduction and sequential data reconstruction. PhD thesis, Dept. of Computer Science, University of Sheffield, UK, faculty.ucmerced.edu/mcarreira-perpinan/papers/phd-thesis.html. Accessed 28 January 2010
35. Brouwer R K (2004) Feed-forward neural network for one-to-many mappings using fuzzy sets. *Neurocomputing* 57:345-360