Neural Comput & Applic (1999)8:290–296

Neural Computing & Applications

A Glossary of Basic Neural Network Terminology for Regression Problems

J. A. Stegemann and N. R. Buenfeld

Department of Civil & Environmental Engineering, Imperial College of Science, Technology and Medicine, London, UK

Complex jargon represents an impediment to newcomers to the field of neural networks. This document presents a glossary of some of the basic terminology that may be encountered in the literature, with an emphasis on the use of multilayer perceptrons and radial basis functions for regression problems.

Keywords: Jargon; Neural networks; Prediction; Regression; Terminology

1. Introduction

Impurities can have complex damaging effects on the hydration of cement, which are important when industrial by-products are used in construction, or when cement is used to treat hazardous waste [1]. For this reason, British Nuclear Fuels Ltd., Euroresiduos S.A., GESENU, and the European Commission have funded a project on 'Neural Network Analysis for Prediction of Interactions in Cement/Waste Systems', based on existing data from the literature, and industrial and academic sources. A database of cement/waste system properties, including cement and waste composition, and final product properties related to durability and leachability, is being assembled by eight partners involved in the field of hazardous waste solidification with cement. This database will include several hundred potential inputs and outputs, which will be used for the purpose of training neural networks constructed by the authors at Imperial College, where neural networks have previously been applied to study concrete durability [2,3] and chloride binding [4]. The authors have found that the existence of complex jargon has hindered their entry into the field of neural network analysis, and have developed the glossary in Table 1 to assist experts from other fields in becoming familiar with basic neural network terminology. The authors' work will focus on use of multilayer perceptrons and radial basis function networks for prediction, and this emphasis is reflected in Table 1. References for the terms are provided where appropriate, but in some instances the authors' understanding stems from the literature in general.

2. Discussion and Conclusions

The terms in Table 1 have been listed in alphabetical order for easy reference, which has the disadvantage that the ordering of the concepts bears no relationship to their use. The following paragraphs attempt to put the concepts in Table 1 in context.

Pattern recognition in data sets such as the properties of cement/waste systems, can be viewed as *function approximation*, which includes both *classification* and *regression*. *Neural networks* are a type of information processing system used for pattern recognition whose *architecture* is based on that of biological nervous systems, and which has a similar ability to *generalise*. Neural networks are composed of a number of *cells*, each of which exchanges information with others in the network through *weighted connections*. In general, the cells are arranged in *layers*, with one or several *hidden layers* of parallel cells lying between the *input* and *output layers*. The weighted inputs received by each cell are summed, including a *bias*, and transferred to

Correspondence and offprint requests to: J. A. Stegemann, Department of Civil & Environmental Engineering, Imperial College of Science, Technology and Medicine, London SW7 2BU, UK.

other cells as a single outcome using a non-linear *activation function*. The connection weights are adjusted by using *training patterns* in iterative *training epochs* (of either *batch* or *sequential learning*), using an appropriate *training algorithm* and *learning rate*, to make the neural network perform the desired function. The knowledge of the trained network is stored in the connection weights as a kind of *distributed memory*, which is said to be *fault tolerant*.

The objective of training is for the neural network to be able to generalise from the training patterns, rather than just being able to reproduce them, so that a correct prediction can be made for a new example. *Overfitting* of the details of the training data, or *overtraining*, can be avoided by using a *validation set* (or *cross-validation*, if data is scarce). The ultimate test of the network is its performance with the *test set*. Regression problems are good candidates for *supervised learning*, in which training patterns are each divided into *input data* and *target outputs*, and iterations are carried out to minimise an *error function* based on the difference between the *computed output* and the target output. However, *unsupervised learning*, in which the training patterns

Table 1. A glossary of basic neural network terminology.

Term	Synonyms	Definition	Reference
1-of-n encoding		encoding for nominal variables whereby each category is assigned a numerical variable, which is encoded as 0 or 1, depending on whether it is absent or present in a particular pattern.	
activation function	excitation function, squashing function, transfer function	a bounded function of infinite domain applied to the weighted and summed inputs to limit the amplitude of the output signal. For multi-layer networks this must be a continuously differentiable function.	adapted from [7] and [9]
architecture	model, paradigm, topology	the arrangement of cells in a neural network. Different architectures vary in the arrangement, type and number of their connections, and in their activation functions, type of learning and training algorithms.	adapted from [9]
batch learning		corrections to the weights are made based on the overall error, which is calculated as the sum of the errors for the individual training patterns.	
bias	intercept, threshold	a weight parameter for an extra input whose activation is permanently set to +1.	[6]
cell	neuron, neurode, node, processing element, unit	a simple linear or nonlinear computing element that accepts one or more inputs, computes a function thereof, and may direct the result to one or more other cells.	[9]
centre	weight	a parameter which defines the location of the centre point of a radial basis function.	
classification	categorisation, labelling	assignment of inputs to discrete output classes.	
cluster	kernel, segment	a region in space containing a relatively high density of datapoints.	
committee of networks	modular network	a combination of several networks which have been trained to solve the same problem.	[6]

Table 1. Continued.

Term	Synonyms	Definition	Reference
competitive learning	winner-take-all	a learning rule where cells compete to respond to an input. The winner then adapts to make itself more like the input.	
computed outputs	dependent variables, responses, predicted values	the outputs predicted by a neural network on the basis of an input vector.	
connections	interconnects, links, synapses	the channels through which information is passed from one cell to another.	
cross-validation	moving window validation, n-fold validation	the available data is divided up into *n* subsets of equal size; one subset is saved for testing, and training and validation are performed using the remaining *n*-1 subsets. Training is repeated *n* times, with each of the *n* subsets acting as the test set once.	
curse of dimensionality		exponential increase in size of input space, and corresponding decrease in data density, as a function of the number of input variables (i.e. dimensions).	
deviation	radius, spread, threshold, width	parameter to define the distance from the centre to the edge of radial basis function.	
distributed memory		the function of a network is stored as weights which are distributed throughout the network.	
epoch	cycle	each repeated entry of the full set of training patterns.	
error-backpropagation	backpropagation, dynamic feedback, learning logic	a method for computing the error gradient, i.e. the derivatives of the error function with respect to the weights, for a feedforward network.	
error function	cost function, objective function	an expression which describes the difference between the computed and target output. Typically the squared error, but sometimes linear error, absolute error or entropy.	
expectation maximisation algorithm		an approach to iterative computation of maximum-likelihood estimates when the observed data is incomplete; used to estimate missing values and in unsupervised learning.	[7] and [6]
fault tolerance	graceful degradation	processing can continue even if some cells or connections are damaged.	
feedforward	forward propagation, static	uni-directional transfer of information.	
function approximation	heteroassociation, prediction, forecasting	the prediction of output values on the basis of input values; includes both classification and regression.	

Table 1. Continued.

Term	Synonyms	Definition	Reference
Gaussian function (normalised)	Gaussian distribution, normal distribution	$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$ where μ is the mean, and σ is the variance.	
generalisation	inference, interpolation, prediction	ability to draw conclusions about highly complex new situations by making associations with previous experience of similar situations.	
gradient descent	steepest descent, standard backpropagation	the iterative changes in the weights during training are proportional to the negative of the first derivative of the total error.	
hyperbolic tangent function	bipolar sigmoid	$f(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}}$ The hyperbolic tangent function is a type of sigmoid function which ranges from −1 to 1.	
input data	stimulus	the vector of variables from which a prediction is intended to be made.	
K-nearest neighbours algorithm		estimation of missing information on the basis of its K-nearest neighbours in the data space.	[5]
layer	field, slab	an arrangement of cells which process information in parallel, i.e., synchronously. Typically one or more *hidden layers* are found between the *input layer* and the *output layer*.	

Note: The convention used in this document is to count the hidden layers of cells; others may count layers of weights, or include the input and/or output layers in the count of layers.

Term	Synonyms	Definition	Reference
learning rate	η	a small parameter which sets the step size for adjustment of weights during training.	
linearly separable		data which can be separated by a straight line in two dimensions, or a hyperplane in n-dimensions.	
logistic function		$f(x) = \frac{1}{1 + e^{-ax}}$ The logistic function is a type of sigmoid function, which ranges from 0 to 1. Usually the slope, $a = 1$.	[7]
minimax transformation		a simple linear transformation to a target range (often 0 to 1).	
multilayer perceptron		a fully connected feedforward error backpropagation neural network with at least one hidden layer.	
neural network	artificial neural network, connectionist model, neural net	a class of flexible nonlinear regression and discriminant models, data reduction models, and nonlinear dynamical systems consisting of an often large number of neurons (i.e., cells) interconnected in often complex ways and often organised into layers.	[9]

Table 1. Continued.

Term	Synonyms	Definition	Reference
overfitting	overlearning	construction or training of a network to fit the details of the training patterns rather than generalise well for new data.	
overtraining		overfitting of the training patterns by continuing to train without the use of an appropriate validation set.	
principal components analysis		projection of data onto the eigenvectors with the largest eigenvalues of the covariance matrix, in order to decrease its dimensionality.	
principal components analysis		projection of data onto the eigenvectors with the largest eigenvalues of the covariance matrix, in order to decrease its dimensionality.	
pruning	growing, brain damage, self-structuring, ontogeny	algorithms for determining the optimal number of hidden layer cells.	
radial basis function		function which involves a distance criterion with respect to a centre, such as a circle, ellipse or Gaussian.	
regression		prediction of the value of a continuous variable y from an input vector x.	[5]
reinforcement learning		during training weights are reinforced for well-performed actions and punished for poorly performed actions.	
sequential learning		weight correction based on the error for each training pattern.	
sigmoid function		a strictly increasing function which exhibits smoothness and asymptotic properties, such as a logistic or hyperbolic tangent function.	[7]
supervised learning		training in which the training patterns are divided into input data and output data, and the number of training iterations is dependent on the difference between the computed outputs and the target outputs for a set of test inputs.	
target outputs	desired outputs, observed values	the output values provided to the network in supervised learning.	
test set		a set of data which the neural network has not previously seen, which is used to test how well the neural network has learned to generalise.	[6]

Table 1. Continued.

Term	Synonyms	Definition	Reference
Note: although the terms are sometimes interchanged, the test set is different from the validation set (see below)			
threshold function	hard-limiting function	noncontinuous activation function, of one of the following types:	
	binary step function, Heaviside function, step function	$f(x) = \begin{cases} 0 \text{ for } x < 0 \\ 1 \text{ for } x \geq 0 \end{cases}$	[7]
	bipolar function, signum (sgn) function	$f(x) = \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ for } x = 0 \\ -1 \text{ for } x < 0 \end{cases}$	[7]
	ramp	$f(x) = \begin{cases} a \text{ for } x \geq a \\ x \text{ for } \lvert x \rvert < a \\ -a \text{ for } x \leq a \end{cases}$	[8]
training	adaptation, estimation, learning, model fitting, optimisation	using examples to adjusting the weights on the connections in the neural network such that the network performs its task correctly. Learning is equivalent to minimisation of an error function.	
training algorithm	learning algorithm, learning rule	the method by which the weights are adjusted during training.	
training patterns	construction sample, example data, training cases, training data	the data set used to train the neural network.	
unsupervised learning	self-organisation	the neural network organises the training data and discovers its emergent collective properties.	[8]
validation set	test data, hold-out sample, cross-validation, verification	a set of data used to test the performance of the network during training, but not used for modifying the weights of the network.	[6]
Note: although the terms are sometimes interchanged, the validation set is different from the test set, and cross-validation is different from ordinary validation (see above)			
weights	parameters, strengths, synaptic weights	the network parameters which are determined by iterative training.	
zero-mean unit-variance transformation	Mean/SD scaling	linear scaling to transform data such that it has a mean of zero and variance of 1.	

are not divided into inputs and target outputs, but rather divided into classes by the network during training, can often be useful in discovering structure in the data. It should be noted that neither *competitive*, nor *reinforcement learning* is particularly applicable to *regression* problems.

The *multilayer perceptron* is probably the most popular network for *regression* applications. The multilayer perceptron is a *feedforward error-backpropagation* network, which uses a *sigmoid* activation function, such as the *hyperbolic tangent* or *logistic function*, as a form of differentiable *threshold function*. Use of one or more hidden layers allows separation of patterns that are not *linearly separable*. Training may be conducted by *gradient descent*, or a variety of more sophisticated algorithms. A number of *pruning* algorithms exist for determining a suitable number of cells for the hidden layer.

Radial basis function networks are feedforward neural networks with a single hidden layer, in which the activation functions for the hidden layer cells are radial basis functions (such as a *Gaussian*). Training of a radial basis function network proceeds in two independent stages. In an unsupervised learning stage, the training patterns are separated into *clusters* (for example, using the *K-nearest neighbours* algorithm). The cluster *centres* and *deviations* are used as the parameters for the radial basis functions in the hidden layer of the network. In the second, supervised, stage of learning, the weights for the connections from the hidden to the output layer are determined.

It has become common to train several neural networks to model a problem, and combine them to form a *committee of networks*.

For most data sets, pre-processing is an important part of neural network analysis. *Principal components analysis* may be used to reduce the number of variables, and avoid the *curse of dimensionality*, missing data may be filled in using the *expectation maximisation algorithm*, and the variables may be transformed, using *1-of-n encoding*, a *minimax transformation* or *a zero-mean unit-variance transformation*.

Tarassenko's *A Guide to Neural Computing Applications* [5] is an invaluable aid for the beginner in neural network analysis and introduces the uses of the concepts summarised here. It is hoped that using this information, and the glossary provided here, the newcomer to the field of neural networks will be able to move on with confidence to the more detailed and valuable information provided in Bishop [6] and Haykin [7], and elsewhere.

References

1. Stegemann JA and Buenfeld NR. Neural network analysis for prediction of interactions in cement/waste systems: introduction. In Mehu J, Keck G and Navarro A (eds), Waste Stabilisation and Environment. Société Alpine de Publications, Grenoble, 1999; 224–331
2. Buenfeld NR and Hassanein NM. Neural networks for predicting the deterioration of concrete structures. In H Jennings et al. (eds), The Modelling of Microstructure and its Potential for Studying Transport Properties and Durability. Kluwer Academic, Dortrecht, 1996; 415–432
3. Buenfeld NR and Hassanein NM. Predicting the life of concrete structures using neural networks. Proceedings of the Institution of Civil Engineers Structures and Buildings, 128, February 1998; 38–48
4. Glass GK, Hassanein NM and Buenfeld NR. Neural network modelling of chloride binding. Magazine of Concrete Research 1997; 49(181): 323–335
5. Tarassenko L. A Guide to Neural Computing Applications. John Wiley & Sons, New York, 1998
6. Bishop CM. Neural Networks for Pattern Recognition. Clarendon Press, 1995
7. Haykin S. Neural Networks – A Comprehensive Foundation. Maxwell Macmillan Canada, Toronto, 1994
8. Simpson PK. Artificial Neural Systems – Foundations, Paradigms, Applications, and Implementations. Pergamon, New York, 1990
9. Sarle WS (ed) Neural Network FAQ, periodic posting to the Usenet newsgroup comp.ai.neural-nets. URL: ftp://ftp.sas.com/pub/neural/FAQ.html, 1998