Revue Roumaine de Chimie, **1998**, *43* (6), 545–551

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. Part 8[1]

THE INFLUENCE OF THE ACTIVATION FUNCTION IN THE ESTIMATION OF THE SENSORIAL SCORES OF RED WINE COLOR

Ovidiu IVANCIUC

Department of Organic Chemistry, Faculty of Industrial Chemistry, University "Politehnica" of Bucharest,
Splaiul Independenţei 313, 77206 Bucharest, Roumania
E-mail: o_ivanciuc@chim.upb.ro

Received September 19, 1996

The sensorial scores of red wine color were estimated with multi-layer feedforward neural networks, using three spectral parameters as independent variables: the dominant wavelength, saturation, and brilliance of the wine sample. A bell-shape activation function, $Act(z) = 1/(1 + z^2)$, for the hidden layer neurons gives good results, both in the calibration and in the cross-validation of the neural network model.

INTRODUCTION

Artificial Neural Networks (ANN) have learning characteristics that allow them to be trained to model nonlinear relationships in data of high dimensionality.[2–4] Such patterns are manifested by significant correlations between molecular structure and physico-chemical properties.[5–15] The ANN model does not require any formulation of rules about the phenomenon under investigation, and forms an internal model by extracting information directly from the properly selected examples forming the training set of patterns.

When computing a Quantitative Structure-Property Relationship (QSPR) model, there is no a priori reason to expect relationships to be linear. Until recently there has been no easy way of exploring nonlinearity in data except by extremely laborious and quite chancy trial-and-error procedures. This has now changed, following the introduction of neural networks.

Recently there has been growing interest in the application of neural networks in the field of QSPR and it has been demonstrated that the neural model provides superior results when compared with the traditional Multi-Linear Regression (MLR) analysis. The key strength of the Multi-Layer Feedforward (MLF) ANNs is that with the presence of hidden layers, neural networks are able to perform nonlinear mapping of the physicochemical parameters to the corresponding property.

The goal of the present paper is to compare the performance of the MLR model and of the MLF ANN model to correlate the subjective judgement of the red wine color with objective parameters (spectral measurements). We will investigate also the use of a new, bell-shape activation function, which we have found to give good results for the calibration and for the cross-validation of the neural model.[13–15]

RESULTS AND DISCUSSION

Recently, several nonlinear regression models were discussed and their performances were compared using a data set from food chemistry.[16] The problem was to model the subjective judgement (sensorial scores) of red wine color from objective measurements (spectral parameters). The patterns set consisted of 58 Barbaresco wine samples characterized by three spectral parameters (independent variables) and a sensorial score of the color, the mean value of the scores from 20 wine experts representing the dependent variable. The three spectral parameters represent the brilliance (x_1), saturation (x_2), and dominant wavelength (x_3) of the wine sample, computed from a spectrum measured between 400 and 700 nm. The 58 patterns are presented in Table 1 and the MLR model for this data set is given by the following equation:

$$y = -48.30(\pm 46.33) - 232.72(\pm 223.26)x_1 + 0.49(\pm 0.47)x_2 + 0.27(\pm 0.26)x_3 \tag{1}$$

$$n = 58 \qquad r = 0.646 \qquad s = 15.306$$

where n is the number of compounds used in the correlation, r is the correlation coefficient and s is the standard deviation. The standard error of estimation of each coefficient at the 95% confidence level is given in parentheses. The partial correlation coefficients are $r(x_1) = -0.644$, $r(x_2) = 0.615$, and $r(x_3) = 0.601$, leading to the conclusion that no single x parameter can estimate with high accuracy the y values. The MLR model is of very low statistical quality, leading to the conclusion that either there is no correlation between the spectral parameters and the sensorial scores, or the relationship is nonlinear.

The collinearity of the three variables is very high, as indicated by the intercorrelation matrix:

	x_1	x_2	x_3
x_2	1.000	–0.935	–0.923
x_3		1.000	0.829

The high intercorrelation coefficients between the spectral parameters are a sign that the three independent variables represent similar information on the characterization of the red wine samples, making them of little use all together in a MLR model.

As already mentioned, the low value for the correlation coefficient in the MLR model has two possible explanations: (*i*) the relationship between the spectral parameters and the sensorial scores of color is strongly nonlinear; (*ii*) not all important wine parameters were identified and used in the MLR model. If the low performance of the MLR model is due to the factor (*i*), the ANN model will give better results, otherwise, in the case (*ii*), the neural model will fail to give a good fit for the sensorial scores of the red wine color.

The ANNs used in the present study are three-layer MLF networks, with three input units representing the three spectral parameters (x_1, x_2, and x_3), and one output unit representing the values of the sensorial scores from Table 1. The training was done with the backpropagation algorithm,[2–4] the input and output scalings were set between –0.9 and 0.9, the initial weights scaling between –0.1 and 0.1, and the momentum was set to 0.8. The usual activation function has a sigmoid shape, but we have investigated the use of a new, bell-shape activation function in the hidden layer, and of a linear function for the output layer. For the bell-shape function we propose the formula: $\mathrm{Act}(z) = 1/(1 + z^2)$, while the sigmoidal function used in the present study was the hyperbolic tangent, tanh, which provides better results than the usual sigmoidal function $\mathrm{Act}(z) = 1/(1 + e^{-z})$.

Kreinovich[17] demonstrated that an arbitrary nonlinear activation function in the hidden layer is sufficient to represent all functions by neural networks. His theorem opened the possibility to use new activation functions in the neural model. In our previous studies[13–15] we have obtained good results with a bell-shape hidden activation function; for the output layer the tanh and linear functions provided better results than the bell-shape function.

Table 1

The food chemistry patterns used in the neural network model; x_1 represents the brilliance, x_2 the saturation, x_3 the dominant wavelength, y the sensorial score of color for the red wine sample under consideration, and y_{NN} the sensorial score computed with the network NN

Sample	x_1	x_2	x_3	y	y_{NN}	$y - y_{NN}$
1	0.099	89.36	606.88	150.8	145.9	4.9
2	0.139	87.15	602.19	122.8	128.0	–5.2
3	0.057	97.74	612.75	166.7	163.2	3.5
4	0.059	98.43	615.14	157.1	151.2	5.9
5	0.068	98.00	609.43	143.0	143.4	–0.4
6	0.107	94.98	607.06	140.1	140.7	–0.6
7	0.101	90.65	607.12	159.7	157.2	2.5
8	0.092	93.38	605.65	150.2	154.3	–4.1
9	0.113	89.81	605.68	152.8	144.9	7.9
10	0.082	96.50	606.33	131.5	141.5	–10.0
11	0.060	96.89	612.08	168.3	165.7	2.6
12	0.194	81.23	598.77	100.1	97.4	2.7
13	0.075	94.62	610.36	160.9	167.4	–6.5
14	0.064	97.82	612.17	153.2	152.1	1.1
15	0.093	94.12	608.57	156.3	153.5	2.8
16	0.047	97.35	616.78	157.4	154.8	2.6
17	0.095	94.15	605.79	123.5	137.5	–14.0
18	0.119	91.02	602.87	134.6	130.0	4.6
19	0.042	98.67	617.30	122.4	122.3	0.1
20	0.122	91.96	604.44	126.3	136.0	–9.7
21	0.087	94.08	608.88	157.4	156.7	0.7
22	0.082	94.60	605.04	130.7	134.1	–3.4
23	0.106	94.80	602.28	128.0	121.5	6.5
24	0.118	90.43	603.76	149.4	145.8	3.6
25	0.116	90.16	602.07	148.8	147.6	1.2
26	0.163	86.35	598.90	121.2	125.7	–4.5
27	0.119	89.59	604.93	143.1	142.4	0.7
28	0.045	97.63	613.28	136.5	137.1	–0.6
29	0.076	96.79	607.82	151.8	146.3	5.5
30	0.072	95.70	611.21	165.3	165.9	–0.6
31	0.082	95.23	607.51	150.9	151.2	–0.3
32	0.056	96.85	611.76	149.7	148.5	1.2
33	0.204	76.30	596.86	88.3	89.5	–1.2
34	0.046	95.83	609.60	127.7	122.7	5.0
35	0.068	97.33	612.01	155.1	153.3	1.8
36	0.091	94.08	607.15	152.0	144.0	8.0
37	0.232	78.66	594.14	71.4	71.5	–0.1
38	0.081	91.58	609.27	146.7	146.6	0.1
39	0.053	94.29	612.78	154.4	150.5	3.9
40	0.085	95.74	609.91	157.4	157.1	0.3
41	0.079	93.29	611.08	157.5	164.4	–6.9
42	0.080	93.74	610.56	165.3	165.1	0.2
43	0.141	83.97	603.80	128.0	130.1	–2.1
44	0.053	97.00	611.39	114.9	134.8	–19.9
45	0.090	94.23	610.49	164.7	161.9	2.8
46	0.113	91.90	606.08	145.3	153.6	–8.3
47	0.083	95.19	609.11	154.7	148.3	6.4
48	0.073	95.40	612.19	165.7	164.9	0.8
49	0.062	96.80	610.44	154.7	149.2	5.5
50	0.073	93.84	611.45	164.7	165.7	–1.0
51	0.046	96.50	615.18	148.9	153.7	–4.8
52	0.068	97.13	612.22	147.1	155.0	–7.9
53	0.099	93.26	607.00	143.0	145.3	–2.3
54	0.042	97.37	615.06	155.7	153.3	2.4
55	0.087	93.86	608.28	159.4	156.0	3.4
56	0.065	95.49	613.58	160.6	162.3	–1.7
57	0.093	96.52	606.11	134.2	128.1	6.1
58	0.040	98.02	619.43	123.3	125.0	–1.7

In order to compare the performance of the ANN model with the statistical results of the MLR equation, we have used the correlation coefficient r and the standard deviation s of the linear correlation between y_{exp} (the experimental sensorial scores) and y_{ANN} (the sensorial data estimated by the ANN model): $y_{exp} = A + By_{ANN}$.

From the topology of the MLF neural networks, the size of the input and output layers of the network is predetermined by the number of input and output variables used in the model, but the number of neurons in the hidden layer must be selected on the basis of empirical trials, in which ANN with different number of hidden neurons are trained to predict the y_{exp} values. The training was done by randomly presenting to the network the values of the parameters for the 58 patterns from Table 1, until the correlation coefficient between y_{exp} and y_{ANN} improved with less than 10^{-4} in 100 epochs.

Selected results obtained with ANNs in estimating the sensorial scores are presented in Table 2. The number of hidden neurons was set between two an five, because a network with one hidden neuron provided a model with a low estimation power, while increasing the number of hidden neurons beyond five did not improve the model.

Since we are interested in selecting the ANN with the best performances, in Table 2 we present the network specifications (the activation function and the learning rate for the hidden and output layers,

Table 2

Calibration results for the estimation of the sensorial scores of red wine color with neural networks. The table reports the network specifications, number of hidden neurons (H), number of training epochs (E), coefficients (A and B) of the linear regression between y_{exp} and y_{ANN}, correlation coefficient (r) and standard deviation (s)

Network Specifications	H	E	A	B	r	s
Act(Hid) = tanh	2	1900	−10.918	1.061	0.872	9.626
Act(Out) = tanh	3	3600	−4.898	1.019	0.910	8.154
η(Hid) = 0.01	4	7700	−4.322	1.011	0.924	7.533
η(Out) = 0.01	5	6900	−4.834	1.016	0.925	7.493
Act(Hid) = tanh	2	2500	−4.383	1.013	0.877	9.440
Act(Out) = linear	3	4400	−0.147	0.985	0.911	8.133
η(Hid) = 0.01	4	3600	−0.303	0.987	0.911	8.124
η(Out) = 0.005	5	9300	−0.611	0.982	0.923	7.557
Act(Hid) = bell	2	300	−8.802	1.047	0.887	9.077
Act(Out) = tanh	3	1500	−6.789	1.038	0.921	7.646
η(Hid) = 0.08	4	1400	−3.750	1.018	0.926	7.438
η(Out) = 0.01	5	8100	−3.182	1.020	0.955	5.861
Act(Hid) = bell	2	4800	0.669	0.992	0.824	11.142
Act(Out) = linear	3	2600	−0.711	0.987	0.909	8.216
η(Hid) = 0.05	4	5100	−1.703	0.991	0.924	7.532
η(Out) = 0.005	5	5000	−2.557	0.998	0.924	7.512

respectively), the number of hidden neurons (H), the number of training epochs (E), and the statistical parameters of the linear regression between y_{exp} and y_{ANN}: the coefficients A and B, the correlation coefficient r and the standard deviation s.

If we compare the predictions of the MLR model represented by eq. (1) with the predictions of the neural networks presented in Table 2, it is clear that the ANN model outperforms regression analysis and provides a better estimation of the sensorial scores. A closer inspection of the results from Table 2 reveals that all neural networks with two hidden neurons offer models with r less than 0.9, with little importance in a quantitative estimation of the sensorial scores. A small improvement is obtained for a size of the hidden layer between three and five neurons, when the correlation coefficient takes values between 0.910 and 0.926. The exception is the ANN with a bell-shape hidden activaton function and tanh output function, with five hidden neurons, which provides the best model, with a higer correlation coefficient, 0.955.

In order to be sure that this high correlation can be reproduced, we have tested a high number of ANNs with this topology, using different initial random weights and different learning rates. In all cases, the correlation coefficient was around 0.960. For example, when the output learning rate, η(Out), was set to 0.005, the learning process was slower, taking 40000 epochs, and provided a neural network denoted by NN with the following performances:

$$y_{\mathrm{exp}} = -0.948 + 1.005 y_{\mathrm{NN}}$$

$$n = 58 \qquad r = 0.961 \qquad s = 5.424$$

The sensorial scores estimated by the network NN, together with the residuals, are presented in Table 1. When compared with the results of the four nonlinear regression methods investigated in ref. 16, the neural method gives always better results, since the correlation coefficient of the nonlinear regression models was lower than 0.9.

Of course, the ANNs performances can be improved by using networks with a higher number of hidden neurons, but in QSPR studies it is very important to take into consideration that feedforward neural networks are universal approximators: they are capable of arbitrarily accurate approximation to arbitrary mappings, provided sufficiently many hidden units are available.[18-20] Recently, the potential of chance correlations in QSAR models was investigated,[21-23] and it was proposed to characterize a network by a structural parameter ρ, which is the ratio of the number of patterns in the training set to the number of connections. The optimum ρ values for the MLF networks lay in the range $1.8 < \rho < 2.2$, and it was reported that networks having ρ values in excess of 2.2 failed to extract relevant features and gave poor predictions. The network NN with five hidden neurons has a ρ value of 2.23, very close to the optimum range. In the present investigation, the use of networks with a higher number of hidden neurons is not justified, since it will lead to ρ values lower than the optimum values.

In order to investigate the predictive character of the ANN model, we have used the Leave-One-Out (LOO) cross-validation method. In the LOO technique, an untrained network is first created. Then one pattern is taken out of the training set of patterns and the network is trained with the remaining patterns. When the learning process is finished, the network predicts the output value for the pattern which was eliminated from the learning set. The pattern is then put back in the set and the next one is taken out to repeat the process, starting with the untrained network. In this manner, each pattern serves as an unknown once and as a training pattern all the other times. The LOO cross-validation results are presented Table 3.

The LOO predictions of the networks provided with the bell-shape hidden activation function are lower than the predictions of the networks with a tanh hidden activation function. The tanh-tanh network with four hidden units has a cross-validation correlation coefficient (r_{cv}) of 0.843, while the maximum r_{cv} for the tanh-linear networks is 0.846, for three hidden neurons. Overall, the LOO predictions of the neural networks are better than the predictions obtained with the set of four nonlinear regression algorithms.[16]

Because the correlation coefficient of the ANN model is higher than that of the MLR model, we can conclude that there is a nonlinear dependence between the spectral parameters and the sensorial scores of the red wine color. Also, because the neural model is not able to offer excellent correlations (with $r > 0.99$), the second conclusion is that there are unidentified wine parameters which influence the sensorial scores, and if one intends to obtain a better model, it is important to use new objective measurements of the wine. As already pointed, the three independent variables are highly intercorrelated, representing similar information on the wine color.

Table 3

Leave–one–out cross–validation results for the prediction of the sensorial scores of red wine color with neural networks. The notations are explained in Table 2

Network Specification	H	A	B	r	s
Act(Hid) = tanh	2	3.990	0.952	0.758	12.846
Act(Out) = tanh	3	5.142	0.946	0.812	11.478
η(Hid) = 0.01	4	1.489	0.968	0.843	10.577
η(Out) = 0.01	5	9.767	0.912	0.837	10.763
Act(Hid) = tanh	2	2.291	0.961	0.787	12.149
Act(Out) = linear	3	−1.043	0.985	0.846	10.499
η(Hid) = 0.01	4	5.901	0.936	0.830	10.966
η(Out) = 0.005	5	7.061	0.927	0.827	11.053
Act(Hid) = bell	2	28.361	0.799	0.725	13.560
Act(Out) = tanh	3	25.887	0.813	0.791	12.041
η(Hid) = 0.08	4	34.678	0.747	0.747	13.089
η(Out) = 0.01	5	32.658	0.765	0.801	11.786
Act(Hid) = bell	2	16.060	0.879	0.680	14.437
Act(Out) = linear	3	10.921	0.902	0.785	12.204
η(Hid) = 0.05	4	28.998	0.782	0.752	12.975
η(Out) = 0.005	5	30.313	0.771	0.738	13.275

The networks with a bell-shape activation function[13–15] in the hidden layer offer the best performances in the calibration phase and one can consider this new type of function as a good alternative to the sigmoidal function when the relationships between input and output patterns are highly nonlinear. Some trials with ANNs having an output bell-shape function offered models with a lower statistical significance.

Neural networks represent an easy way of exploring nonlinearity in data, with superior results when compared with a set of four nonlinear regression algorithms[16] which were used to model the subjective judgement (sensorial scores) of red wine color from objective measurements (spectral parameters).

ACKNOWLEDGEMENT. We acknowledge the partial financial support of this research by the Ministry of Research and Technology under Grant 310 TA69 and by the Ministry of National Education under Grant 7001 T34.

REFERENCES

1. Part 7: O. Ivanciuc, *Rev. Roum. Chim.*, in press.
2. D. E. Rumelhart and J. L. McClelland, "Parallel Distributed Processing", MIT Press, Cambridge, MA, 1986.
3. D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature,* **1986**, *323*, 533–536.
4. P. D. Wasserman, "Neural Computing", Van Nostrand Reinhold, New York, 1989.
5. B. J. Wythoff, *Chemom. Intell. Lab. Syst.*, **1993**, *18*, 115–155.
6. J. Gasteiger and J. Zupan, *Angew. Chem. Int. Ed. Engl.*, **1993**, *32*, 503–527.
7. J. R. M. Smits, W. J. Melssen, L. M. C. Buydens and G. Kateman, *Chemom. Intell. Lab. Syst.*, **1994**, *22*, 165–189.
8. N. Bodor, M.-J. Huang and A. Harget, *J. Mol. Struct. (Theochem)*, **1994**, *309*, 259–266.
9. A. A. Gakh, E. G. Gakh, B. G. Sumpter and D. W. Noid, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 832–839.
10. L. M. Egolf, M. D. Wessel and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 947–956.
11. M. D. Wessel and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 68.

12. O. Ivanciuc, *Rev. Roum. Chim.*, **1995**, *40*, 567–574.
13. O. Ivanciuc, *Rev. Roum. Chim.*, **1995**, *40*, 1093–1101.
14. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye and J. P. Doucet, *J. Chem. Inf. Comput Sci.*, **1996**, *36*, 644–653.
15. O. Ivanciuc, J.-P. Rabine and D. Cabrol-Bass, *J. Chem. Inf. Comput Sci.*, **1997**, *37*, 587–598.
16. I.E. Frank, *Chemom. Intell. Lab. Syst.*, **1995**, *27*, 1–19.
17. V. Kreinovich, *Neural Networks,* **1991**, *4*, 381–384.
18. G. Cybenko, *Math. Control Signals Syst.*, **1989**, *2*, 303–314.
19. K. Funahashi, *Neural Networks,* **1989**, *2*, 183–192.
20. K. Hornik, M. Stinchcombe and H. White, *Neural Networks,* **1989**, *2*, 359–366.
21. T. A. Andrea and H. Kalayeh, *J. Med. Chem.*, **1991**, *34*, 2824–2836.
22. D. T. Manallack and D. J. Livingstone, *Med. Chem. Res.*, **1992**, *2*, 181–190.
23. D. J. Livingstone and D. T. Manallack, *J. Med. Chem.*, **1993**, *36*, 1295–1297.