Zeitschrift: Mitteilungen aus dem Gebiete der Lebensmitteluntersuchung und Hygiene = Travaux de chimie alimentaire et d'hygiène

Herausgeber: Bundesamt für Gesundheit

Band: 89 (1998)

Heft: 1

Artikel: Comparison of multivariate statistics and neural networks as classificaiton tools in the field of wine analysis

Autor: Kaufmann, Anton

DOI: https://doi.org/10.5169/seals-983136

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 23.08.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Mitt. Gebiete Lebensm. Hyg. **89**, 97–117 (1998)
Eingegangen 8. Mai 1997. Angenommen 22. Januar 1998

Comparison of Multivariate Statistics and Neural Networks as Classification Tools in the Field of Wine Analysis

Key words: Multivariate statistics, Neural networks, Wine, Chemometrics

Anton Kaufmann
Schuler-Weine, Seewen SZ

Currently working at:
Official Food Control Laboratory of the Kanton Zürich (Kantonales Labor), Zürich

Introduction

Some of the most demanding work performed in food laboratories are classification tasks: Is a flavour compound of natural or artificial origin? Does a certain wine originate from the declared geographical origin? Does an expensive olive oil contain cheaper vegetal oils? Such problems can seldom be solved by looking at a single chemical component. More often the analyst is compelled to evaluate the concentration of several analytes (variables) in order to derive an answer. The most simple approach is a digital multivariate approach. If a sample contains more than a specified amount of substance «a» and less than another specified amount of substance «b» etc. then it belongs to the unalturated group. Such a rather crude digital approach does extract only a small amount of the present information, because it neglects the interaction of the individual variables.

Such hidden information can be unscrambled by methods belonging to the family of multivariate statistics (linear discriminant analysis, principle component analysis and partial least square etc.). The mathematical background (matrix algebra) of these methods had been developed long before computers capable to perform such extensive calculation became available. Multivariate statistics is based on a solid mathematical background and is therefore thoroughly understood. However, because of the complex mathematics involved, multivariate statistics has not yet became popular among non-statisticians.

Neural networks on the other hand are a rather new development. Their inventions have been inspired by the growing performance of computers. While most classical multivariate methods are straight forward oriented methods, neural networks are based on iteration processes. In other words they utilise a trial and

error approach. Neural networks are most commonly viewed as black-boxes. Since a thorough understanding of the underlying principles is not compelling, neural networks have become popular among many different users. The black-box stigma, however, is probably responsible why neural networks are not readily accepted by many statisticians. Researchers using neural networks, however, claim that their methods imitate the problem solving approach used by the human brain. They state correctly that human thinking does not follow the path of classical statistical analysis. Because of this two distinguished school of thought, classification problems are often solved by the method that the researcher is most inclined to and less by the approach which is most suitable for the particular problem.

There are two basically different concepts within multivariate statistics and neural techniques. Either a supervised or an unsupervised approach is used. Supervised methods (e.g. linear discriminant analysis, back-percolation neural networks) require a corresponding answer for each sample. Hence, if the researcher is interested in the geographical origin of wine, he uses the measured concentrations of different substances (variables) and the known geographical origin (answer) of the given wine as input. Unsupervised methods (e.g. factor analysis, kohonan neural net) do only require the different substance concentrations (variables). They do not need a corresponding answer. Hence the algorithm «classifies» the samples in a way which is obvious from the variable point of view. The resulting discrimination plot might however show a different discrimination than expected by the researcher. The algorithm might for example rather plot red and white wines apart than spread the samples according to the geographical origin (1). Because of this characteristic, questions like the determination of the geographical origin of wine are rather solved by supervised methods. Supervised methods are more powerful when it comes to detecting minor differences, however, if not the necessarily care is applied, over-optimistic predictions result.

The present paper discusses the classification performance of multivariate statistical and neural network methods. The focus is on the practical problems related to food laboratories applications (geographical origin of wine) and less on underlying theoretical concepts. The size of the reference data, a key-factor to avoid over-optimistic predictions, is thoroughly discussed.

Multivariate statistics

There are several different classification approaches available. In the field of multivariate statistics, the most widely used methods are different cluster techniques, linear discriminant analysis, SIMCA, Partial Least Square and Principle Component Regression. Several authors have discussed the use of these techniques in the field of wine analysis (2–5). This paper discusses only the use of Linear Discriminant Analysis (LD). Previous work (1, 6, 7) showed that LD is well suited for the used data set. The mathematical background as well as possibilities and limitations have been discussed in a previous paper (1). LD analysis belongs to the family of supervised methods (see above) and is based on multilinear-regression,

hence LD utilises only linear relationship. Therefore, non-linear structures cannot be uncovered. However, if non-linear relationships are suspected, a non-linear transformation of the variables (e.g. logarithmation) can be applied. By utilising untransformed and transformed variables, non-linear models can be built (1).

Neural networks

There are actually several, very different neural network techniques used. Nevertheless, many researchers use the term neural networks synonymous with the back-propagation algorithm. A survey (8) indicated that 80% of all chemical oriented neural network papers utilise the back-propagation technique. Still popular are Kohonan nets which permit Graphical Multidimensional Scaling. Linear learning machines, Hopfield nets, ABA memory and Counter Propagation are of rather historical interest. A more recent development is the Back-Percolation (BPR) algorithm. Neural networks were used successfully to determine the geographical origin of Italian olive oil by utilising concentrations of different fatty acids (9).

Modern Multilayer networks (counter-propagation, back-propagation and back-percolation BPR etc.) consist of an input layer, one or several hidden layers and one output layer. Each layer contains several neurons (circles). Figure 1 represents a 5-9-3 net. It accepts 5 independent inputs (e.g. 5 analytical variables) and produces 3 different outputs (e.g. three possible geographical origins). The 9 neurons in the hidden layer are the intermediary between the input and the output neurons. In a standard net, all neurons from the input layer are connected (lines) to all the neurons in the hidden layer. Hence, any hidden layer neuron shown in figure 1 is fed by inputs from each input layer neuron. The hidden layer neurons

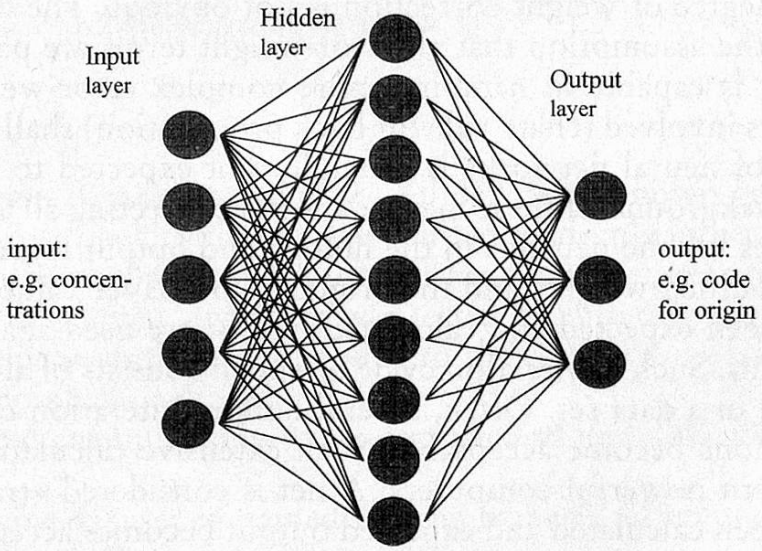

Fig. 1. Architecture of a Multilayer Neural Network: The depicted net accepts 5 inputs (e.g. 5 different analytical parameters) and calculates the 3 outputs (e.g. probability for belonging to one of these three possible geographical origins). This is done by the use of 9 neurons in the hidden layer. The value of a particular neuron is determined by summing the contribution of all neurons from the previous layer.

are again connected to all the neurons in the output layer. The neurons in the input layer are variables which contain the input variables (e.g. concentration of a certain analyte), while the neurons of the output layer represent the output values (e.g. a digital variable representing certain geographical origins). Each connection between the different neurons is represented by a weight. The weight is a number which indicates the importance of a particular connection. Hence, a large weight indicates that the neuron on the left side of the connection makes an important contribution to the value of the neuron located on the right side of the connection. The value of a neuron in the hidden or the output layer is calculated by summing the different inputs. This is done by multiplying the value of a preceding neuron by the weight associated to the connection. The added products produce the value of the neuron. Hence the neurons in the hidden layer of figure 1 are calculated by summing the products of the five inputs. The information of a neural network is located in the weights. The untrained algorithm starts with randomly determined weights. By summing the products of weights and corresponding neurons in the input layer, the values of the hidden layer neurons can be calculated. Knowing these values, the same approach is used to determine the values of the output level neurons. Obviously such a calculated output is likely to differ from the desired output, since the initial weights consist of randomised numbers (untrained net). It is the aim of the training process to obtain output values, equal as closely as possible to the desired output. This is done by adjusting the weights. Hence, the important following step is the backward proceeding weight correction. Each output neuron corrects now their individual incoming weights in order to produce a smaller deviation between calculated and expected output. This is not very difficult for the neurons in the output layer, because the deviations from the desired output are known. However, since the correct values of the neurons in the hidden layer are unknown, the degree of weight correction is not obvious. The back propagation algorithm uses the assumption that all error-weight terms are parabolic. BPR on the other hand, is capable in handling more complex error-weight curves. The intricate calculus involved (chain rule for back propagation) shall not be discussed here. The user of neural networks is generally not expected to comprehend the mathematical background of these methods. After correcting all the weights backwards, the values for the neurons in the hidden and output layer are recalculated by using the modified weights and the original input layer values. The decreased deviations between expected and calculated output are used again to correct the preceding weights. Such an iteration cycle or epoch consists of all samples (e.g. all available wines) in a data set. Often, several hundred iteration cycles are needed until the deviations become acceptable. Such extensive calculations can only be solved by modern powerful computers. A net is considered «trained» when the difference between calculated and expected output becomes acceptable. Using the input from an unknown data-set (e.g. wines from unknown origin), the trained neural network produces hopefully the correct classification.

Experimental

The leading Swiss Wine Magazine and an International Swiss Daily organise biannually a wine tasting championship. Participants receive 18 neutral numbered bottles containing particular wines. The participant is asked to recognize the wine samples correctly. He is aided by a supplied guidebook which describes the wines and the regions they are originating from. The samples included wines from eight countries and three different continents.

It was attempted to participate in this championship by utilising analytical data and multivariate statistics only. The wines were analyzed by ion-exclusion LC and gradient ion chromatography, AAS and wet Chemistry. 18 different analytical parameters (carbohydrates, organic and inorganic anions, lead, cadmium and pH) were available (7). Our reference data, obtained during a period of several years, contained more than 6000 wines from practically any geographical origin.

The results of this investigation have been submitted for publication (7) before the dead line of the championship. The calculated classification of all 18 wines was found to be correct. However, some classifications were associated with a rather high degree of statistical uncertainty. Therefore the mentioned set of data was evaluated again by the use of neural networks (BPR).

Analytical data

16 variables were applied for both methods: trehalose, glucose, fructose, glycerin, ethanol, acetic acid, lactic acid, chloride, galacturonic acid, succinic acid, malic acid, tartaric acid, sulfate, phosphate, citric acid and pH. No attempts to eliminate unimportant variables were made. All analytical data (variables) were standardized (set to average = 0 and standard deviation = 1)

Groups

Three different data sets were used. All data with known geographical origin came from our data bank. This data was splitted randomly into a training and a test data set. The training set was used as to teach the neural network or the discriminant approach. The developed algorithm was tested by inputing the test data. Set-ups which proved successfully when applied to the test set were used to classify the unknown championship data set.

The presented championship wines consisted of three blocks: Rosé, white and red wines.

The white wine data set contained samples from six different countries/regions:

origin of wine	group
Waadt, Switzerland	a
Fendant, Wallis, Switzerland	b
Chardonnay, USA	c
Johannisberg, Wallis, Switzerland	d

Valencia, Spain	e
Bordeaux, France	f

The red wine data set contained samples from nine different countries/regions:

origin of wine	group
Dôle, Wallis, Switzerland	j
Burgundy, France	k
Châteauneuf-du-Pape, France	l
Rioja, Spain	m
Cabernet Sauvignon, Chile	n
Bordeaux, France	o
Chianti classico, Italy	p
Barbera, Italy	q
Ribero del Duero, Spain	r

Training/Test set size

The available reference data contained 650 white wines. 319 wines were assigned to the complete training set and the remaining 331 to the test set. The 962 red wines were divided into 461 wines to the complete training set, the test set contained the remaining 501 wines. The reliability of the prediction depends on the number of wines available for the training set. In order to observe this effect, the number of samples in the training set was varied. Besides the «complete» training set (319; 461 samples), a «halve» training set (160; 230 samples) and a «quarter» training set (80; 115 samples) were produced. Hence the number of wines in the quarter training set is four times lesser than the complete training set. The number of samples in the test set remained unchanged.

Multivariate statistics

The applied statistical technique was LD. The used software was «Statistica 5» (StatSoft)

No elimination of variables was made.

Neural networks

The BPR algorithm was applied. The used software was «Braincell 3» (Promised Land Technologies). For efficient classification, the number of net output neurons was made equal to the number of different wine groups. Groups were coded digitally (1 resp. 0). The number of hidden neurons was set according to software defaults.

White wine 16-20-6 net. Red wines 16-24-9 net. Learing rate: 0.1 Initial weights: +– 0.4. The number of cycles (number of iterations) was set to 40, 80, 160 and 320. For each condition the net was retrained five times. The reported classification accuracy was produced by averaging these five repetitions.

Results

Classification of test set samples (known wines)

Effect of number of cycles

The statistical method of LD does not perform iterations, hence this parameter is only relevant to the neural network approach. At the end of a cycle, all net weights are adjusted in order to decrease the training error. The more cycles (iterations) are performed, the smaller the training error becomes. However, obviously the researcher is more interested in the test data error than in the training data error. Figure 2 represents the performance for white wines while figure 3 shows the classification of the red wines.

Generally the training set error is clearly smaller than the test set error. The more cycles performed, the smaller the training error becomes.

Effect of training data size

The size of the training set was reduced. As evident from figure 2 and 3, the reduction of samples in the training set increases the number of wrong classifications in the test set. The present data do not permit a statement whether BPR or LD analysis is more sensitive towards a «thin» training set. An increase in numbers of cycles performed, reduces the test set classification error. However, the error curve rises again when training is based on the halve or quarter training set. Obviously an over-training of the net occurs the earlier the less samples the training set contains.

Comparing the LD algorithm to BPR, BPR performs in all three white wine training sets better than LD. LD excelled only in one red wine training set (quarter) BPR.

Unlike BPR, LD issued warnings when processing the «thin» quarter training set. The «tolerance» warnings are a feature of LD which should warn the uses from relying to strongly on the present «thin» training set.

Furthermore, the software refuses any calculation if one training group contains a variable with no variance.

Training error versus test error

Depending on the network set-up and the number of cycles, the training error can be reduced to a very small level. As mentioned above, such an overtrained net will perform poorer on a test data set. LD analysis does also show a smaller training error than a test error. However, the discrepancy is much smaller than in BPR. The size of the test error related to the size of the training data set seems not to be the same for both algorithms. LD training data error decreases with a shrinking training set and at the same time the test error increases. Considering neural networks, the training data error seems to increase when the training data size is decreased. This is true if we compare nets with the same number of cycles. However, the number of weight corrections is proportional to the product of cycles and number of

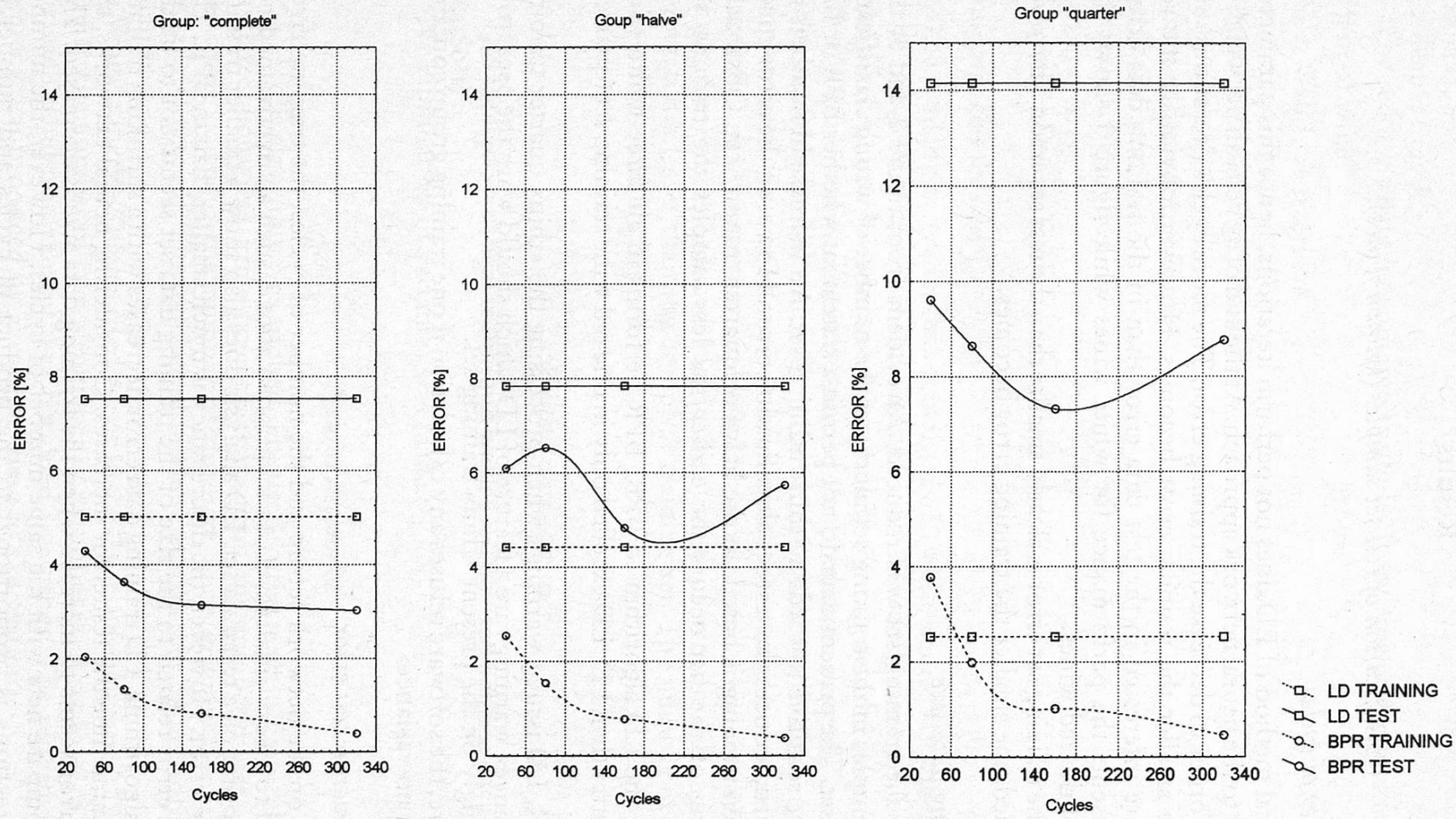

Fig. 2. An algorithm should not only predict known samples (training set) but also unknown samples (test set). The average classification error for white wines originating from six different geographical origins is shown. Reducing the available samples in the training set (complete, halve, quarter) strongly affects the classification errors. The training set error decreases, however more important is the fact that the test set error (unknown samples) increases. Good results can only be obtained, if sufficient reference samples are available. In the case of neural networks, the number of iteration should be carefully optimized.

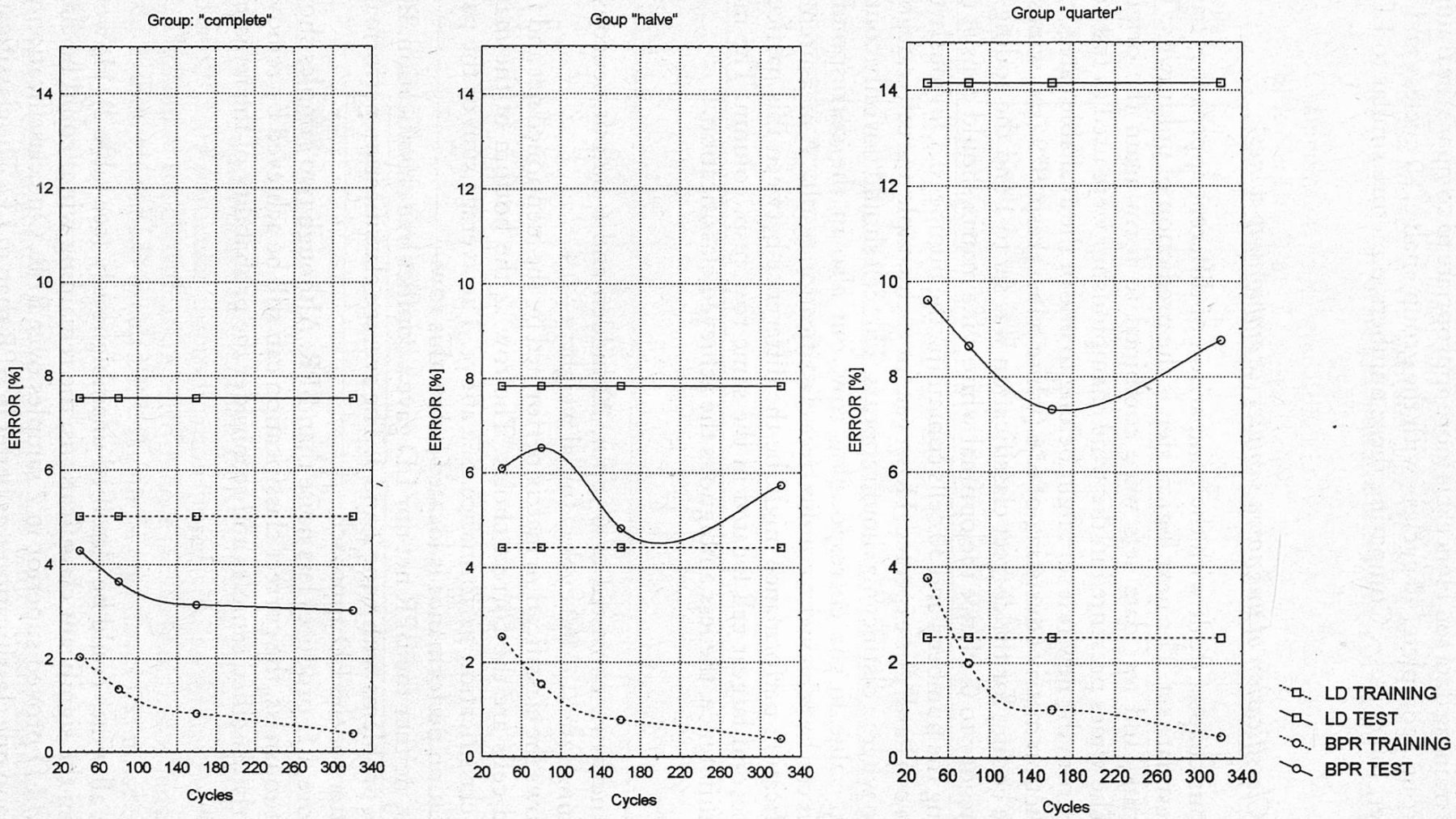

Fig. 3. Red wines originating from nine different geographical origins were classified. The figure is otherwise equivalent to figure 2. Again, using less training samples (complete, halve and quarter) increases the classification error of unknown test samples. There is no clear difference between neural networks and discriminant analysis. Neural networks build by using limited training samples «quarter» should not be overtrained by performing too many iterations.

samples in the training data set. Hence it is more appropriate to compare the training error of the group «complete, 40 cycles» with the group «halve, 80 cycles» and the group «quarter, 160 cycles». Comparing these numbers, a picture similar to LD is observed.

Classification of unknown samples (championship wines)

The previous chapter dealt with the comparison of different approaches based on a known test data set. The next logical step is the prediction of unknowns. The previous training and test data sets were combined to form again the original training set. The wines presented in the wine championship were used as test set.

Each wine from the new test set might belong to one of the available categories. Hence a prediction matrix consisting of 6*6 (white wines) resp. 9*9 (red wines) is obtained. The coding for the correct classification was set to 1 and the coding for the wrong category to 0. Hence the optimal white wine matrix should consist of 6 cells containing the number 1 and 30 cells containing the number 0. Obviously this ideal situation can hardly be expected. However, for each sample one cell (category) should exist to where a distinct maximum appears. This cell should not only contain the maximum value in the corresponding row, but also in the corresponding column. If this can be achieved, a conflict-free classification results. A measurement for the classification performance might be the difference between this particular cell and the second largest cell located in the same row resp. column. The more positive this difference, the less ambiguous the achieved classification.

White wines

The BPR network consisted of 6 input, 20 hidden and 6 output neurons (3 layers). Learning rate: 0.1. 160 cycles. Initial weight: +– 0.4

Table 1 shows the classification matrix performed by the mentioned net and LD. The italicized cells are the correct choice. The row at the bottom of the matrix indicates the classification performance. See above. The average sum of the particular classification performance is located below this row.

As evident, neither the BPR net nor LD gave a conflict free classification. Both methods have problems with wine 2 and 3. The average classification performance is slightly better for the LD approach.

Furthermore, LD produces less noise than BPR. Although two samples show a certain ambiguousness, the correct classification can still be achieved by removing the most clearly classified samples and reconsider the probability of the remaining samples again.

Red wines

Again not all wines could be classified unambiguously (see table 2). However, after removing the significant classifications, the remaining wines could be classified correctly. LD proved superior in 7 samples over BPR. One classification was almost equal and one classification was better in BPR than in LD. The classification

Table 1. Predicting of the unknown white championship wines: Italicized cells indicate correct answer. A «1» indicates a perfect fit, a «0» a complete misfit. The row «classification performance» indicates the clearness of a particular prediction.

Championship white wine BPR/LD
BPR Net (16/20/6) 160 cycles

Groups:	a	b	c	d	e	f
Wine:						
1	–0.25	0.13	–0.13	0.04	*1.20*	0.02
2	0.89	*0.13*	–0.03	0.00	0.01	0.00
3	0.52	–0.06	0.12	*0.43*	0.01	–0.02
4	0.08	–0.05	–0.03	0.09	0.11	*0.80*
5	0.06	–0.07	*1.14*	–0.06	0.02	–0.10
6	*0.97*	–0.01	–0.03	0.04	0.05	–0.02
	classification performance:					
	0.08	–0.75	1.02	–0.09	1.07	0.69
	sum:		2.02			

LD

Groups:	a	b	c	d	e	f
Wine:						
1	0.00	0.00	0.00	0.00	*1.00*	0.00
2	0.78	*0.22*	0.00	0.00	0.00	0.00
3	0.72	0.00	0.00	*0.27*	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	*1.00*
5	0.00	0.00	*1.00*	0.00	0.00	0.00
6	*0.94*	0.06	0.00	0.00	0.00	0.00
	classification performance:					
	0.16	–0.56	1.00	–0.45	1.00	1.00
	sum:		2.15			

where LD showed poorer results than BPR concerns group j and o (Wallis and Bordeaux). This particular discrimination posed similar difficulties when using BPR. The sum of the classification performance is clearly higher when applying the LD algorithm. Hence LD excels BPR. Again LD produces less noise than BPR.

Practical problems

An important practical difference between multivariate statistics and neural networks is the «fuzzyness» of neural outputs. LD is a mathematical method which

Table 2. Predicting of the unknown red championship wines: Italicized cells indicate correct answer. A «1» indicates a perfect fit, a «0» a complete misfit. The row «classification performance» indicates the clearness of a particular prediction.

Championship red wine BPR/LD
BPR Net (16/24/9) 160 cycles

Groups: / Wine:	j	k	l	m	n	o	p	q	r
1	0.58	–0.16	–0.06	0.04	–0.15	*0.91*	0.00	–0.10	–0.07
2	–0.01	–0.23	0.25	–0.03	–0.04	0.48	–0.15	*0.74*	–0.01
3	–0.02	0.32	*0.08*	0.01	0.12	0.24	0.00	0.15	0.09
4	*0.68*	0.20	–0.10	0.05	0.06	–0.11	0.01	0.31	–0.10
5	–0.02	0.47	–0.03	0.09	0.01	–0.10	–0.12	–0.17	*0.89*
6	0.01	0.13	–0.02	*0.64*	0.08	–0.09	–0.01	0.21	0.05
7	0.00	*0.46*	0.02	0.02	0.09	0.06	0.17	0.21	–0.03
8	0.02	0.04	–0.14	0.11	*0.60*	0.50	–0.01	–0.08	–0.04
9	0.01	0.21	–0.03	0.02	0.00	–0.04	*0.73*	0.12	–0.02
	classification performance:								
	0.11	–0.01	–0.24	0.44	0.11	0.33	0.52	0.26	0.42
	sum:		1.93						

LD

Groups: / Wine:	j	k	l	m	n	o	p	q	r
1	0.75	0.01	0.00	0.00	0.00	*0.24*	0.00	0.00	0.00
2	0.00	0.02	0.01	0.03	0.00	0.11	0.00	*0.68*	0.15
3	0.00	0.35	*0.07*	0.00	0.13	0.15	0.00	0.25	0.04
4	*0.99*	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*1.00*
6	0.00	0.01	0.00	*0.88*	0.00	0.00	0.01	0.08	0.02
7	0.00	*0.87*	0.01	0.00	0.00	0.00	0.03	0.08	0.00
8	0.00	0.00	0.00	0.00	*0.83*	0.17	0.00	0.00	0.00
9	0.00	0.13	0.01	0.00	0.00	0.00	*0.78*	0.08	0.00
	classification performance:								
	0.24	0.52	–0.28	0.80	0.66	–0.50	0.65	0.43	0.85
	sum:		3.37						

delivers exactly repeatable answers. Neural networks, however, show a certain fuzzyness. The reason for this can be traced to the random numbers applied to the initial weights of the net. Even several hundreds of cycles (weight corrections) do

not lead to a complete convergence. There is not much variance observed within an output of an «easy» discriminant problem. Focusing, however, on difficult classification problems, the variability of the output increases strongly. Figure 4 depicts the outputs of 10 consecutive trainings. It represents the output values for the correct wine (white wine championship set). A perfect net would produce the code 1 for each output. Because of this output variance, it is advisable to repeat classification of test data by the use of retrained nets. All previous presented data were, therefore, calculated as the average of five retraining sessions.

Data sets of equal size are definitely faster processed by LD than by BPR. Considering the retraining time and the involved human interactions, LD deliver much faster results than neural networks. Neural networks offer several parameters which can be modified in order to improve performance. Considering the inheritant output variance, it is very time consuming to modify intelligently these parameters. Some neural network software offer self optimization algorithms. Such features test different network setups independently of any human interaction. At the end, an «optimum» net is proposed. Using the red wine data set and a 120 MHz Pentium computer, such a calculation took us 2 days.

LD issues warnings if certain criteria are not met. E.g. too high correlation among variables or lack of variance in a certain variable. Some of these warnings, which indicate highly instable models, can be overruled by the user. However, if a

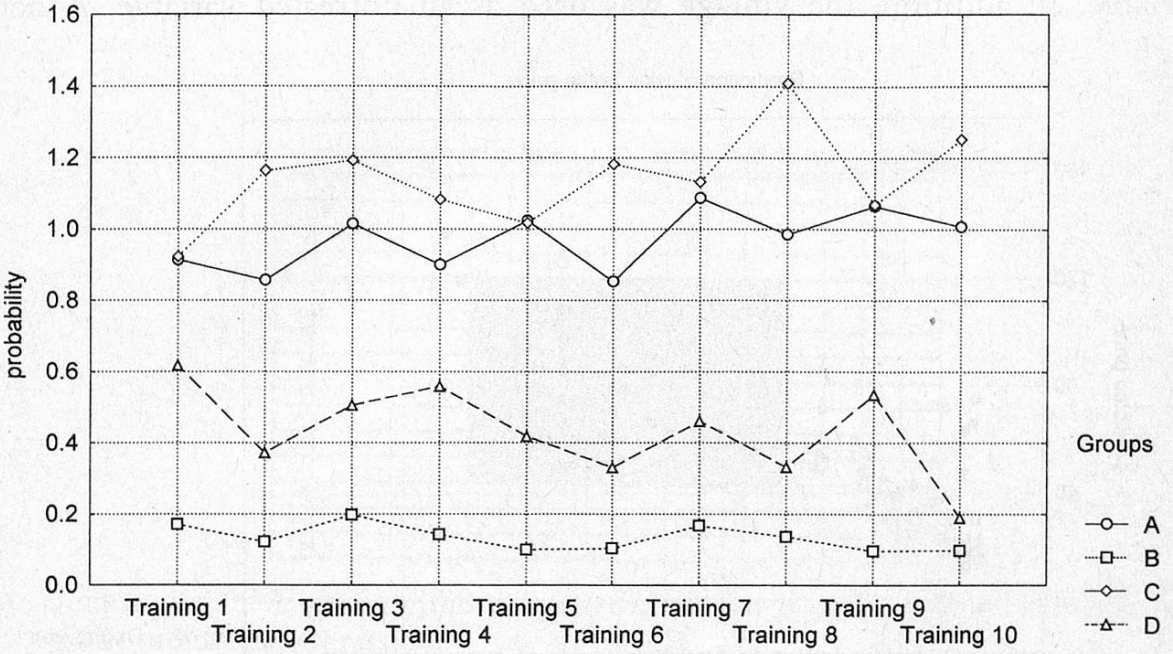

Fig. 4. Repeating a neural network training with the same data produces a slightly different output. The fuzziness of the output obtained by 10 times repeated training sessions is shown. Four white wines from the championship samples were attempted to classify. A perfect fit is indicated by an output of «1», a misfit is represented by a «0». Therefore the samples belonging to group «A» and «C» can be clearly identified. On the other hand, samples from group «B» and «D» are poorly recognized. However, the fuzziness of the output makes decisions not easier. Because of these variations, figure 2 and 3 indicate the average output consisting of 5 training sessions.

group with no variance is detected in any of the variables, the program will not continue with the calculation. No such warnings or «refusals» are issued when using standard neural networks.

Structural information versus noise

The experiments above showed that the two data sets contain a high proportion of geography related variance (structural information). It is not evident whether the remaining variance can be called error. This remaining portion might contain structural information not found in one data set or structural information not interpretable by the algorithm. The poorer classification of the test set compared to the training set is due to noise which was modulated by the used model. The more flexible a model, the higher the likelihood that noise is mistaken for structural information. An experiment was therefore designed to measure the «ability» of the two methods to model a rather doubtful relationship. The task to be performed was the prediction of the wine bottle price based on the wine name and the vintage. This is definitely an overstretching of any statistical or neural technique, since the individual letters in the wine name do not carry interpretable information. 40 Bordeaux wines were selected from a wine retailer catalog. Each letter in the wine name was converted into the corresponding ASCI code and used as an individual variable. In addition, the vintage was used as an untreated variable. A neural

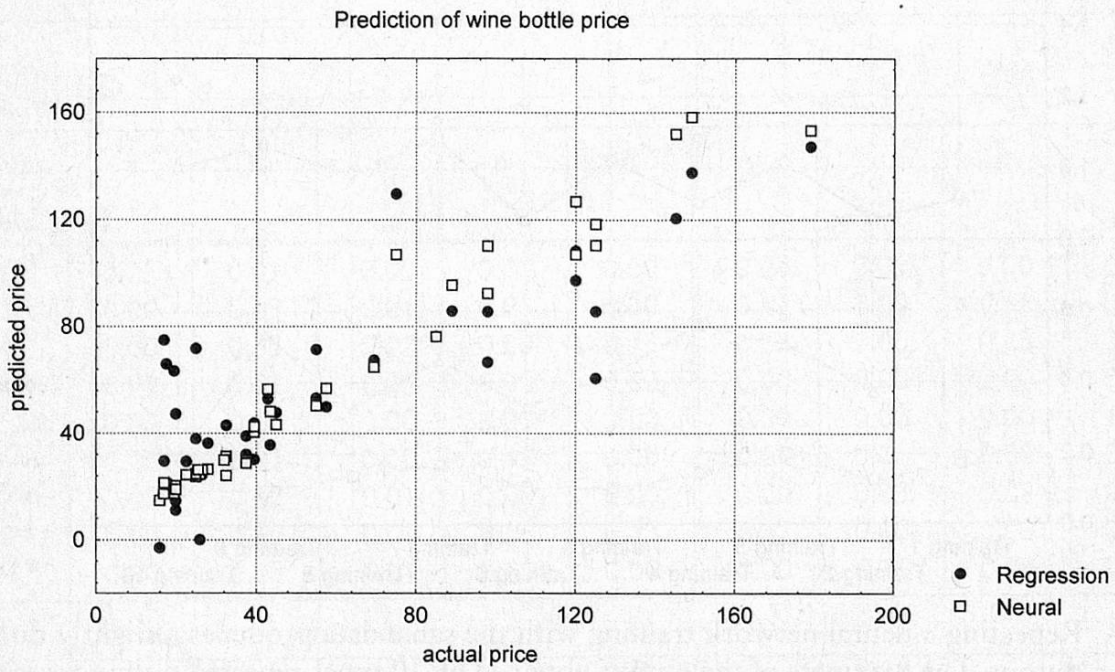

Fig. 5. «Garbage in garbage out» is known to be a basic law. Hence one might attempt to predict the price of a wine by studying the individual letters of a wine name. The retail price of several Bordeaux wines and their corresponding wine names (ASCI codes) were used as input. Such a «nonsense» data set can still produce «beautiful» results. This example indicates the danger present, if a researcher uses an insufficient number of samples. Multilinear regression was clearly more robust than the neural network.

network resp. multilinear regression MLR was used to predict the bottle price. Figure 5 shows a plot of the expected versus the observed wine price. As visible, BPR produced clearly smaller deviations from the regression line than MLR.

The correlation produced by MLR was statistically significant. The main contributing variable was, however, the vintage. This variable was negatively correlated with the wine price. After removing this plausible variable, the remaining variables produced a regression line with a 0.05 significance. Unlike BPR, MLR estimated a negative price for two wines, indicating a highly doubtful prediction capability.

Unique patterns

Figure 6 shows a (x, y) matrix. The value represented by an individual cell is either 0 or 1 (indicated by the colors black and white. The resulting figure is easily recognized by a human as a square within a square. The (x,y) values and the color coding (0/1) can be processed by LD and BPR. A BPR net with sufficient neurons in the hidden layer detected the structure and showed a very small error when tested by a test data set. LD, however, fails completely in recognizing the square. This effect is not surprising since the figure was produced in such a way that there is no correlation between the x, y variables and the color coding.

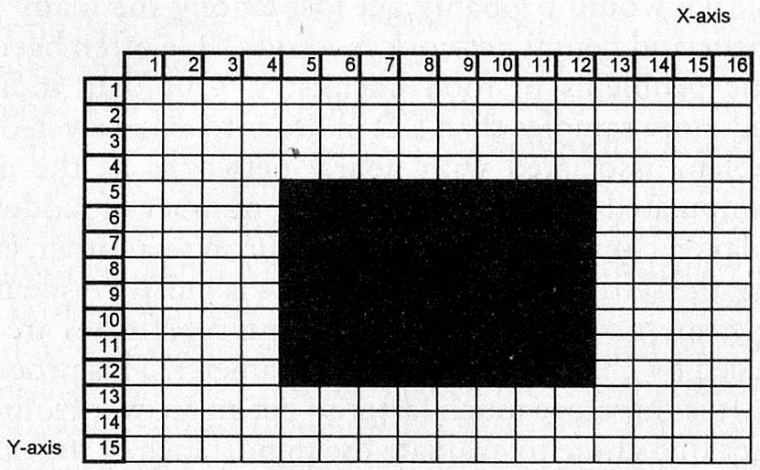

Fig. 6. Unique pattern: A square within a square can be recognized by a neural network, but not by multilinear regression.

Discussion

Working with multivariate statistics or neural networks is something very different. Modern software makes multivariate statistics available to the non-statistician. However, the output created by such programs often fills several screen pages. These detailed test statistics and diagnostics provide valuable information,

yet, they are likely to scare off many users with only a passing interest in the method. The need to acquire a thorough understanding of the underlying mathematics is stressed by most statisticians. Hence, most users of multivariate statistic are aware of the basic pitfalls. This is probably one reason why multivariate statistics has not encountered the same degree of popularity, but also discredit, as neural networks had. The principle of neural networks on the other hand, is much easier to understand than multivariate statistics. This statement however does not refer to the intricate calculus involved. Today's powerful PC make neural networks available to almost anyone. If the user settles with the proposed default values, he will soon receive an answer to his problem. Furthermore, neural networks produce clearly readable results, outputs without the numerous diagnostic tools known from multivariate statistics. A user with only a partial understanding of his data and unfamiliarity with neural networks might, however, not be armed against potentially wrong answers. Unlike most statistical methods, neural networks do not issue warnings if certain conditions (e.g. insufficient number of samples or lacking variance) are violated. A researcher ventures on dangerous ground if he uses more analytical variables than he possesses samples per group. It is probably a human trait to believe that knowledge increases by knowing more variables from the same sample. However, the more variables we consider, the more samples we need to have. For example, it does not help to study the distribution of the different notes (variables) in four pieces (samples) of classical music in order to find out which two pieces were written by Beethoven and which by Mozart. Our common sense tells us that we would probably get lost among the many notes. However, multivariate statistic and neural network methods have often been abused to solve similar unrealistic problems in food chemistry. Our data indicate that neural networks demand more samples than LD in order to avoid over-optimistic results.

A major problem associated with neural networks is, the many parameters which can be manipulated (cycles, learning rate, number of hidden layers, number of nodes in each layer, connections from one neuron to another, error cap etc.). As mentioned earlier, the setting of these parameters is mostly done by trial and error. Because of the fuzziness of the net output, several repetitions are needed to figure out the effect caused by changing one initial parameter. This process becomes very time consuming. It is therefore much faster to get an answer from LD. Performing LD, the researcher finds time to evaluate the importance of the different variables. The elimination of noise containing variables, leads to more stable models.

Unlike LD, neural networks are not bound to linear relationships, hence such a net is capable to model practically any given set of data. Since many relationships observed in nature are non-linear, the use of such algorithms might be considered desirable. However, the capability of neural nets to describe such models is the very reason, why they might perform poorer in some other samples. The additional degrees of freedom available in neural networks increase the likelihood that an irrelevant random noise is recognized as structure which can be modeled. LD on the other hand is bound to strictly linear terms. This limitation prevents the construction of too sophisticated models. A technique like neural networks extracts such non linear information. However, the problem arises when a distinction

between information relevant variance and error variance should be made. The more flexible the algorithm, the more noise will be wrongly considered as information which can be modeled by the algorithm. Hence LD is likely to extract too little, while neural networks extract too much variance. The noise extracted by neural network might therefore overweight the additional non-linear structures found. Hence a poorer classification could result. Such a trained algorithm will show a small training error, but a much larger test set error. This was clearly shown by the nonsense prediction of the bottle price based on the wine name. Not the model of non-linear algorithm itself is of questionable value. The problem is rather the lack of additional information indicating the significance of the output. In other words, the researcher does not know whether the used algorithm might have produced the same result by supplying only random numbers instead of the data set. This lack of «quality control» is a major disadvantage of neural nets when compared to multivariate statistics. It is therefore very important to test every neural network training output with a test data set. Otherwise some very misleading conclusions might be drawn.

Neural networks perform best when each group in the training set contains a similar number of samples. Large differences among the group sizes cause under- or over-training for a particular group. Dealing with a real world problem, equal numbers of samples in each group are rarely encountered. In our case, the training data set contains many more wines from Bordeaux than Ribera del Duero. One approach to deal with this problem is the duplication of such a «weak» group in the training range. There are indications that the output improves, however, this method is rather questionable since it does not introduce new variance.

While the neural approach proved superior in the test data set (originating from our own data bank), the championship samples were classified better with discriminant analysis. Because there are only 18 championship samples, this difference is probably not statistically significant. However, there are reasons which might explain these differences. The training and the test data set contain wines which originate from our data bank. The two groups were created by splitting this data bank randomly into two equally large groups. However, both groups contain only wines purchased by our company or wine samples which were evaluated but not purchased because of various reasons. Hence the data present in our data bank cannot be considered completely unbiased. The screening of samples performed by the wine-tasting team, and the analytical laboratory, affect the selection of the samples to a certain degree. As a winery laboratory, many of our samples are analyzed before the wine is properly aged and ready for consumption. The wines presented for the championship consisted, however, mainly of wines at the peak of ripeness. The championship wines were all analyzed in one series, which was obviously not the case for the training set wines. Hence there are several factors which could be responsible for minor differences between our wine data bank and the wines presented in the wine tasting championship. If this is true, it would not be difficult to explain, why the neural network performed better in the test data set than in the actual championship data set. Then it could be concluded that the answer provided by the network did not only reflect the geographical origin but also to a

minor degree the typical characteristics of the wines being analyzed in our laboratory. Therefore, an important conclusion from this paper is the recognition of differences existing between samples collected by different sources. For a delicate problem like the geographical origin of wine, it is advisable to include samples collected from completely different sources: A friend bringing a wine from his France holiday. Wines intended for sauce production or wines produced by amateur wine makers. Such samples present a formidable test to any model.

Teaching multivariate statistics has focused very strongly on comprehending the matrix algebra needed to perform a calculation. Such a mechanic approach has scared off many interested biologists and chemists. There has been a highly academic discussion about the quality of the data needed to perform multivariate methods. However, these strict requirements are hardly met by any real-world data. Still such realworld problems can be successfully solved by multivariate methods, if a sufficient number of samples is present. Unfortunately, the statistical literature has neglected to issue such practical guidelines. There is generally the assumption of multivariate normal distribution, a requirement which is rarely encountered: If a researcher is interested in differentiating milk from different animal species, he might utilize the measured fatty acid concentration and multivariate statistics. There is certainly a Gaussian curve for each fatty acid when considering healthy lowland cows. It is however unlikely that cows with a metabolic disorder or cows feed in an elevated mountain meadow, produce a milk which is perfectly described by these standard distribution curves. If the researcher receives his first milk sample from such a «mountain cow» multivariate statistics or neural networks might probably consider it as lama milk. Hence, knowing one's data is of uttermost importance, it is definitely more important than knowing how to appreciate the different techniques used to perform the delicate operation of matrix inversion.

Although there are some critical aspects concerning neural networks, such algorithms present another, independent view on the available data set. The recognition of the unique structure «square within a square» does indicate this. This example might support the claim made by neural network specialists that neural network imitate more the way of human thinking than statistical analysis. Therefore neural networks are a valuable tool, a tool which demands a responsible application. It seems to be very advisable to combine the two techniques in order to avoid some pitfalls. Classification problems should be best addressed first by factor analysis. This unguided method uncovers groupings which are actually present in the data set. The clustering produced might, however, not correspond with the groups the user is looking for. The user might be interested in differentiating wines from two very similar Swiss wine growing regions. Factor analysis would tell him that this particular question is difficult to answer, but that there are clear groups present, corresponding to red, white and rosé wine. In order to answer the original geographical question, the user is advised to divide the wines into these three groups before proceeding to the geographical related question. By this way the dominating color related variance is removed from the data set. Factor analysis is also a very useful tool to eliminate extreme outliners which likely distort further calculations. The now more homogeneous groups can be processed by LD or neural

networks. As mentioned above, neural networks should always be verified by a test data set. Performing LD «tolerance» levels should be watched and unimportant variables removed by backward elimination. If there is a good correspondence between the neural and the statistical approach, the researcher can be highly confident of the quality of his results.

Conclusion

Although this work does only cover the classification of wine, the applied approach is valuable to almost any other food product (identity and origin of oil, water, fruit juice etc.). It is likely that classification techniques work even better for such problems. The reason for this could be attributed to the individuality of wine. Most food products are expected to show neither batch to batch nor year to year variation. Wine is something different: Vintage related differences are not rejected, but rather expected by many consumers. The ambition of most wine-makers is to produce a product as individual and distinguished as possible. Hence there is a lot of individual variance superimposed on the geography related variance.

Multivariate methods (statistics and neural networks) have been increasingly employed in recent years. The advent of a new technique is always met with over-optimistic supporters on one side and skeptical critiques on the other side. There have been plenty of papers where answers to the most complex classification problems have been proposed. Not seldom a minimum of samples but a formidable number of analytical parameters (variables) were utilized. A test of such models was often not included, or it was promised as a future paper (which probably never appeared). While most scientific papers demand a high standard, e.g. calibration curve, recovery at different levels, statistical evaluation etc., chemometric papers lacking any verification of the proposed models, however, are still too often accepted. There were as well as there are questionable results obtained by multivariate methods. This should, however, not be an argument against a technique as a whole. Multivariate techniques are presently at a similar stage as capillary GC was 20 years ago. There were only expensive and unreliable columns available, solvent effect were not understood. Fluctuating split ratios and analyte discrimination were strong arguments against capillary GC. While capillary GC has nowadays practically replaced packed GC, multivariate methods are not expected to obtain the same importance. However, such algorithms will probably increasingly be used to extract more information from the available analytical hardware. They might produce faster and therefore cheaper results. Multivariate methods might also be very successfully used for screening purposes. Chemometrics is definitely not a solution to any analytical problem. Its responsible use, however, can solve certain difficult tasks which are otherwise impossible or simply too expensive to elucidate.

Summary

The classification performance of linear discriminant analysis has been compared to back-percolation neural networks. The focus was on the determination of the geographical origin of grape wine. Neural networks showed lower classification errors for the available training data set than the statistical method (linear discriminant analysis). However, a different picture resulted when unknown samples from a wine tasting championship had to be recognized. The linear discriminant analysis approach performed clearly better than the neural network approach.

Zusammenfassung

Die Leistungsfähigkeit von linearer Diskriminanzanalyse und back-percolation neuralen Netzwerken wurde verglichen. Die Aufgabenstellung war die Bestimmung der geographischen Herkunft von Wein. Neurale Netzwerke zeigten bei Trainingsdatensätzen kleinere Klassifikationsfehler als die statistische Diskriminanzanalyse. Ein anderes Bild resultierte bei der Erkennung von unbekannten Weinproben aus einer Degustationsmeisterschaft. Die lineare Diskriminanzanalyse produzierte deutlich klarere Klassifikationen, als dies bei neuralen Netzwerken beobachtbar war.

Résumé

En vue d'une classification des vins, l'efficacité de l'analyse discriminante linéaire a été comparée avec la méthode de la «back-percolation» de réseaux neuraux. La tâche était la vérification de l'origine géographique du vin. Le réseau neural donne une meilleure classification des échantillons connus. Cependant, ce n'était pas le cas avec les échantillons qui ont été utilisés dans un concours officiel de dégustation de vin. Dans ce cas, la classification obtenue avec l'analyse discriminante linéaire a été plus claire.

Literature

1. *Kaufmann, A.:* Multivariate statistics as a classification tool in the food laboratory. J. Assoc. Off. Anal. Chem. **80,** 665–675 (1997).
2. *Etiévan, P.* and *Schlich, P.:* Varietal and geographic classification of French red wines in term of major acids. J. Sci. Food Agric. **46,** 421–438 (1989).
3. *Martin, P.* and *Zhang, B.:* The use of trace element data to complement stable isotope methods in the characterization of grape must. A. J. Enol. Vitic. **45,** 79–85 (1994).
4. *Eder, R.* and *Wendelin, S.:* Klassifizierung von Rotweinsorten mittels Anthocyananalyse. Mitt. Klosterneuburg **44,** 201–212 (1994).
5. *Moret, I.* and *Scarponi, T.:* Caratterizzazione del vino D.O.C. «Soave classico» mediante determinazione di componenti dell'aroma e applicazione del metodo chemometrico «SIMCA». Riv. Vitic. Enol. Conegliano **4,** 254–262 (1985).
6. *Kaufmann, A.:* Ansätze zur Bestimmung der Weinprovinienz mittels Routineanalysendaten. Mitt. Gebiete Lebensm. Hyg. **85,** 566–578 (1994).
7. *Kaufmann, A.:* Prognostischer Einsatz von multivariaten Techniken in der Weinanalytik. Mitt. Gebiete Lebensm. Hyg. **87,** 604–622 (1996).

8. *Zupan, J.* and *Gasteiger, J.:* Neural networks: A new methods for solving chemical problems or just a passing phase? Anal. Chim. Acta. **248,** 1–30 (1991).
9. *Forina, M.* and *Armanino, C.:* Eigenvector projection and slmplified non-linear mapping of fatty acid content of italien olive oils. Ann. Chim. (Rome) **72,** 127–143 (1982).

Anton Kaufmann
Kantonales Laboratorium
Postfach
CH-8030 Zürich