USER'S MANUAL
Part 3:
Examples

Table of Contents - Part 3: Examples

Introduction

This part of manual includes practical examples from many different areas of science. Every example begins with a general introduction to the problem and the presentation of the goal. This is followed by the modeling section, where we explain how to model the phenomenon, which tools are used and what we expect to obtain. The model should be verified, and for this reason we determine the value of penalty coefficient to use, how to obtain it, and why to use that particular value. Comments which give additional information and additional notes, may help users in modeling their specific phenomenon. Because the presented examples are shown at different levels, some of these subsections are not strictly in the order described above.

Example 1

1. GENERAL PROBLEM: Time series

1.1 General

One of the central tasks of empirical science is to build models on the basis of experimental data. A practical example of such a problem is the modeling of a time series. We recognize quickly, that the modeling of a time series is not only a typical technical problem, but is also a characteristic of many other real life disciplines, e.g. stock- and gold-market prediction. A recent research in a chaotic time series showed that neural networks exceed conventional linear and polynomial predictive methods by many orders of magnitude. We will not relate here with a chaotic time series; we will give a simple illustrative sinusoidal example. It will be shown, that a simple example can help us to understand the phenomenon and help to avoid difficulties, which might appear in more complicated cases.

1.2 Modeling of the Phenomenon

A time series can be described mathematically as

$$Y = f(X), \text{ where } X=X_0, X_1, X_2, ..., X_N$$

or in vector form as

$$\{Y\}^T = \{Y_0, Y_1, Y_2, ..., Y_N\}^T$$

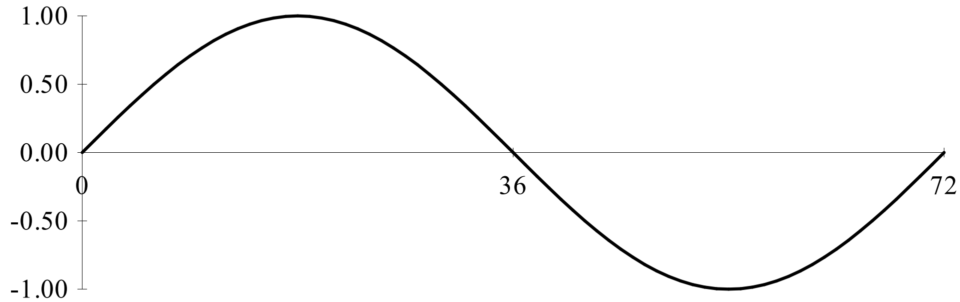


Figure 1.1. Sinusoidal function, described with 72 model vectors.

The variables depends on time. The description seems more natural, if we write

$$\{Y(t)\}^T = \{Y(t_0), Y(t_1), Y(t_2), ..., Y(t_N) \}^T$$

This is the definition of the problem. The three sequential sinusoidal values are used as the input data to predict the fourth, the output. The same description is used as many times as necessary to completely describe one sinusoidal period. Each model vector may be written in general form as:

$$\{\mathbf{mv}\}^T = \{Y(t_1), Y(t_2), Y(t_3); Y(t_4)\}^T$$

1.3 Verification of the model

The first step in a verification of the model was the determination of the optimal value of penalty coefficient using the verification tool. After several trials, the value 0.003 was found as the optimal value. Then we started with a prediction, where in each sequential step previously predicted values were used. This procedure can be written in general form as:

1. step: feed sin(t-2), sin(t-1), sin(t-0) to the network to compute sin(t+1),
2. step: feed sin(t-1), sin(t-0), sin(t+1) to the network to compute sin(t+2),
3. step: feed sin(t-0), sin(t+1), sin(t+2) to the network to compute sin(t+3),
4. step: continue with the same procedure for as many steps you want!

Figure 1.2 shows the results of the prediction for three periods. As can be seen, the results are somehow unnatural - the model in unable to predict descending near the top (maximum values) and climbing near the bottom (minimum values) of the curve.

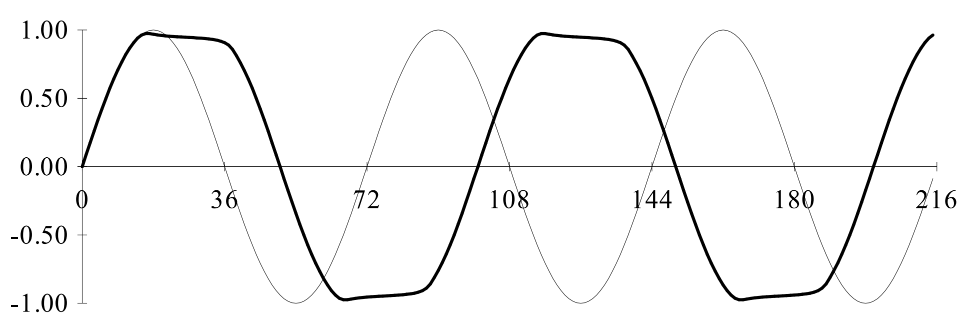


Figure 1.2: Prediction of sinusoidal time series at penalty coefficient = 0.003.

While the solution was not what we wanted, we repeated the prediction again, this time with a lower value of penalty coefficient. The results, which were better, are shown in the Figure 1.3.

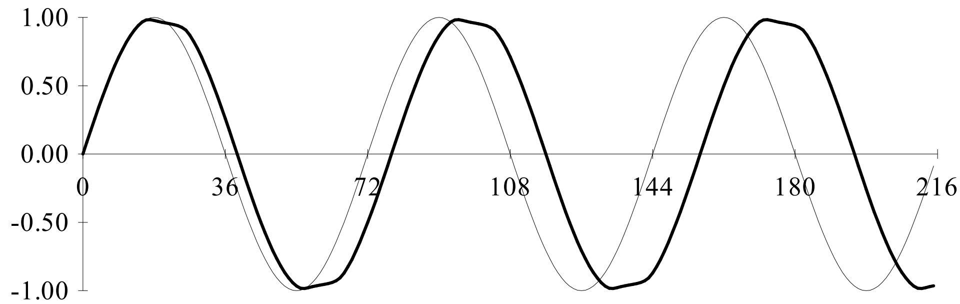


Figure 1.3: Prediction of sinusoidal time series at penalty coefficient = 0.002.

The results from Figure 1.3 suggest a very low value of penalty coefficient. After the prediction has been repeated a third time with penalty coefficient = 0.001, the predicted values were almost exact. We obtained a true sinusoidal curve, except very near the top and the bottom of the curve again; which normally can not be seen in the graphic presentation.

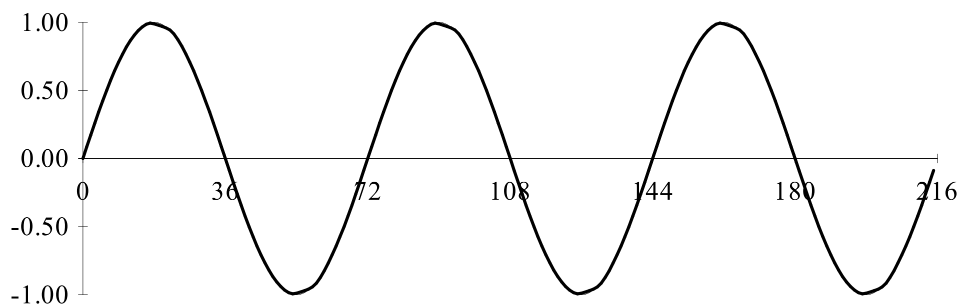


Figure 1.4: Prediction of sinusoidal time series at very low penalty coefficient (0.001).

Let us see the results in the case of a higher penalty coefficient, for example: at 0.01. This situation, which can be seen as a result of many classical neural networks, is shown in Figure 1.5. As it was mentioned in the User's Manual, Part 2, the penalty coefficient is strongly correlated with the learning error in the BP NN. A greater allowed learning error (learning threshold) corresponds to the a higher penalty coefficient. In the situation where the learning threshold is set too high, the BP NN can not learn the phenomena sufficiently. A time series is very sensitive to disturbances in the data, so the BP NN should be trained perfectly for the acceptable predictions. This corresponds to a very low penalty coefficient.

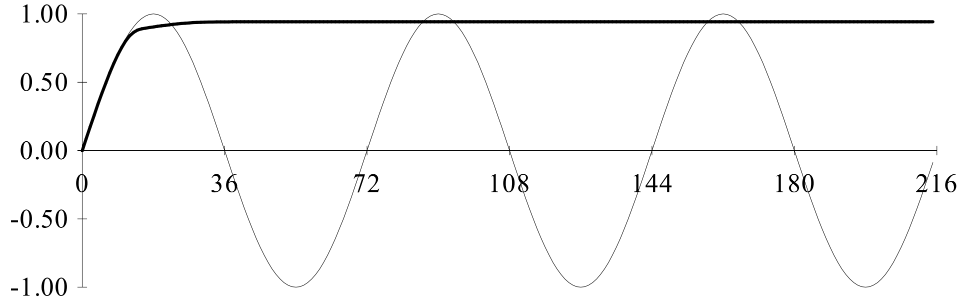


Figure 1.5: Prediction of sinusoidal time series at penalty coefficient = 0.01.

NOTE: Input data (learning set) are stored in the data base, named EXAMPLE1.CSV. The predicted results, shown in Figures 1.2, 1.3, and 1.4 are stored in the data base EXAMPLE1.RES. Users may check these values by simulating the above described procedure by typing manually the predicted values in aiNet prediction window.

1.4 Comments

Users may try to use the example with 360 discrete points. They will find that more data will not solve the problem. Only very high precision data (small value of penalty coefficient) will give the right results. This illustrates two very important facts:

- the problem is very sensitive to small disturbances (prediction errors); prediction of time series might show unusual behavior in more complicated cases.
- a lot of data cannot guarantee a good prediction. For example, take weather forecasting. Taking into account every molecule around the earth, we could predict (theoretically) the weather exactly. But such a huge amount of data demands very high precision, which can be lost through the computation process. Therefore it can be concluded, that there exists not only the optimal value of penalty coefficient, but also the optimal ratio between variables of the phenomenon, and an acceptable prediction, according to the available (software and hardware) tools.

Example 2

2. GENERAL PROBLEM: Approximation of Non-linear Functions

2.1 General

Neural networks and neural network-like systems are approximators already by their nature [1]. For this reason, they may be used for the approximation of non-linear functions. The use of aiNet will be shown using a simple example of the previous day-temperature example (in User's Manual, Part 2). The influence of penalty coefficient is shown first. This is followed by the selection of variables (model vector preparation) in different situations.

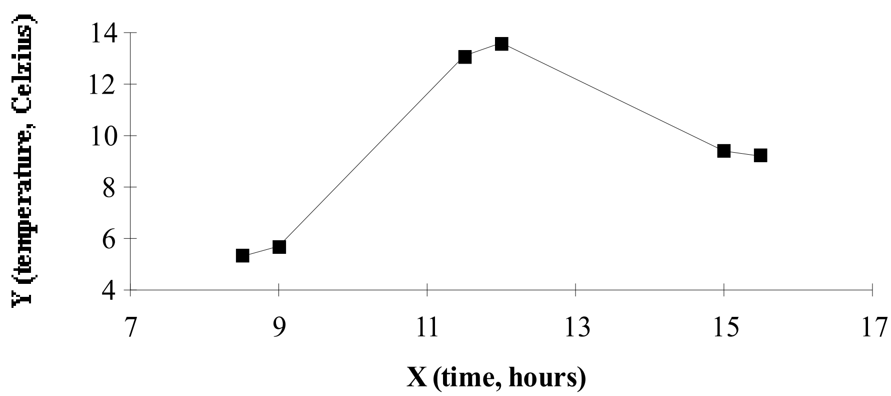


Figure 2.1: The phenomenon of the day-temperature changing, described by six measurements.

The phenomenon of the day-temperature changing is described using six measurements. Mathematically, this may be presented using six model vectors. Time points are shown in the horizontal direction (usually labeled as X), while day temperatures are shown in the vertical direction (usually labeled as Y).

[1] Carpenter, W., C. and Barthelemy, J., F., **Common misconceptions about neural networks as approximators**, Proceeding of the 3rd International Conference on the Application of Artificial Intelligence to Civil and Structural Engineering, in Neural Networks and Combinatorial Optimization in Civil and Structural Engineering (Topping, B. H. V., editor), pp. 11-18, 1993.

2.2 Modeling of the Phenomenon

We are interested in the day temperature as a function of time, which, mathematically appears as:

$$T = T(t)$$

Figure 2.2 shows solutions of the above equation. Results are obtained by using the prediction tool (see aiNet: *Prediction*) in discrete time points, starting at 7:30 with a step length of 6 minutes. It should be noted that aiNet gives different solutions depending on the value of penalty coefficient.

Figure 2.2./a/ clearly shows that a very low value of penalty coefficient gives a very rough curve. In comparison to the classical BP neural network, this condition corresponds to a state, where the neural network memorizes learning patterns very well (model vectors from the data base, according to our convention). A generalization of such acquired neural networks is usually inadequate, while networks learn model vectors very well.

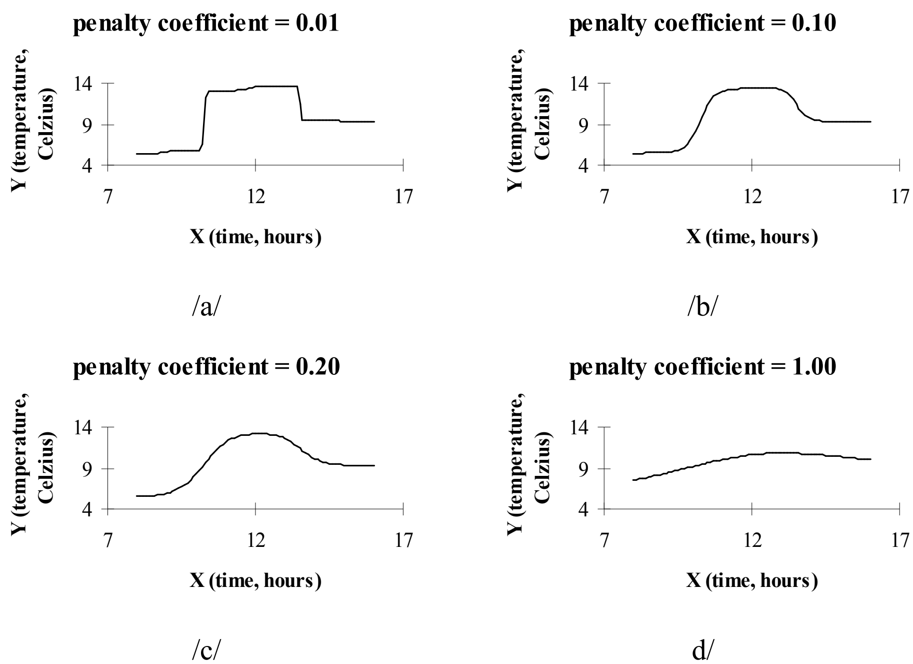


Figure 2.2: Graphical presentation of modeling the day-temperature changing phenomenon as a function of penalty coefficient.

Large values of penalty coefficient give a very smooth curve. Increasing the value of penalty coefficient further, leads to a constant value.

Figures 2.2./b/ and 2.2./c/ show two intermediate states. It is evident that the best solution is obtained by using the penalty coefficient value of 0.20 (see Figure 2.2./c/).

Suitability (validation) of the aiNet model may be assessed visually. This is limited to very simple cases. In multidimensional cases, which are the cases in real life, suitability may be assessed on the basis of various measurements, given by aiNet (see User's Manual, Part 2, Chapter 3). The assessment of *RMS error on variable* or/and *total RMS error* is usually used. In the above example, we tried to determine the penalty coefficient value according to the total RMS error. It should be noted, that a verification tool is used for this purpose.

Now, examine the inverse problem. The above example gives the solution of an equation T = T(t). How is the inverse mapping obtained?

$$t = t(T)$$

For example, we want to know, when some appointed temperature was reached? While the function, which describes the phenomena, is not one-to-one, we have to deal with the problem as it is described in the User's Manual, Part 2 (see Chapter 1 and Chapter 2). Model vectors should be instead of

model vector	=	time	temperature
mv_1	=	8.50	5.30
mv_2	=	9.00	5.70
mv_3	=	11.50	13.10
mv_4	=	12.00	13.60
mv_5	=	15.00	9.40
mv_6	=	15.50	9.20

be written as

model vector	=	time	temperature	period
mv_1	=	8.50	5.30	1
mv_2	=	9.00	5.70	1
mv_3	=	11.50	13.10	1
mv_4	=	12.00	13.60	2
mv_5	=	15.00	9.40	2
mv_6	=	15.50	9.20	2

The third discrete variable describes the ascending part of the curve (morning) when it has value "1", and descending part of the curve (afternoon) when it has value "2". In the mathematical sense, the description of the phenomenon is unchanged.

There are now two input variables: temperature and period (part of the day), when the temperature is the object of interest (morning [1], afternoon [2]). An output variable is the time point in which the temperature has been given value.

NOTE: There are two data bases. The first one EXAMPL2A.CSV is described with only two variables (one input and one output). The second, EXAMPL2B.CSV is described with three variables (two inputs and one output). The whole phenomenon is described by six model vectors. Results from Figure 2.2 are obtained on the basis of input data, stored in EXAMPL2A.PRD.

2.3 Verification of the Model

While the shown example is of illustrative nature, the model is not verified by available mathematical tools, provided by aiNet. However, we can conclude very quickly that the model gives visually acceptable results at value 0.20 for the penalty coefficient (see Figure 2.2./c/). The answers to the given questions in case of using the third (discrete) variable, are exact at the same value of penalty coefficient, too.

2.4 Comments

If the mathematical description of the phenomenon is not using the third (discrete) variable, the results will represent just average values at times when an appointed temperature is reached. Such results do not have any practical value or applicable meaning. We need to be careful in modeling the phenomenon, which should be described in such a way, that inverse mapping,, if it is required, may be obtained without problems.

If aiNet predicts the time in the morning when the temperature is 8°C, it produces an answer: at 8 o'clock and 55 minutes. The value of penalty coefficient was 0.15. One of interesting thing that should be mentioned is that aiNet can tell us also, when the morning or afternoon is. In this case the time is the input variable.

Example 3

3. SENSOR PROCESSING: Character Recognition

3.1 General

Unlike mathematics and science, which, respectively, pursue pure beauty and an understanding of nature, neurocomputing technology pursues the development of useful things. From the earliest days of neurocomputing, neural networks have been recognized as being exceptionally well suited for solving the problems in sensor processing. While neural networks can handle a large amount of information at once, they have opened doors to interesting applications in the area of pattern recognition - the process of visually interpreting and classifying symbols. It should be mentioned that sensor processing primarily means pattern recognition, although other functions such as filtering and data compression are also important application areas. [[2]]

The problem of an automatically recognizing characters in printed or hand-written material has been studied for decades (for example, see [[3]] and [[4]]). A number of technical and market successes have been achieved in character recognition. The United States Postal Service uses character recognizers extensively for reading printed addresses on letters, banks utilize character recognition machines for reading the numerical amounts of checks etc.

In character recognition, there are two important performance measures: *acceptance rate* and *substitution rate*. The acceptance rate is a number of characters per million (in per cent) that the system accepts as readable. The substitution rate is a number of characters per million (in per cent) accepted characters that the system classifies incorrectly. Both of these systems rates are measured by testing the system on huge volumes of text drawn at random from the environment in which the system is expected to operate. The Best systems achieved an acceptance rate of approximately 95%, and a substitution rate of less than 5%.

A simple example of character recognition is shown here, primarily to illustrate how to use aiNet in such cases. The problem may be defined as a classification of single numerals written by hand. A pixel feature description, without pre-processing is used in the example, although no serious

---

[2] Hecht-Nielsen, R., **Neurocomputing**, Addison-Wesley Publishing Company, Inc., 1990.

[3] Guyon, L. et al, **Comparing different neural network architectures for classifying handwriting digits**, Proc. of the Int. Joint Conf. on Neural Networks, **II**, pp. 127-132, IEEE Press, New York, June 1989.

[4] Hecht-Nielsen, R., **Nearest matched filter classification of spatiotemporal patterns**, Applied Optics, **26** (10), pp. 1892-1899, 1987.

researcher suggests such a description as being appropriate for character recognition. We will show two different ways of coding the pixel information and the obtained results in both cases.

3.2 Modeling

Signs were constructed in a box which is divided into 7 columns and 7 rows, as shown in Figure 3.1. The recognition system should be able to distinguish three different signs, which represent numbers 0, 1 and 2. Figure 3.1 shows the numbers presented with 12 different symbols (four samples of each number). Accordingly, the data base consists of 12 model vectors.

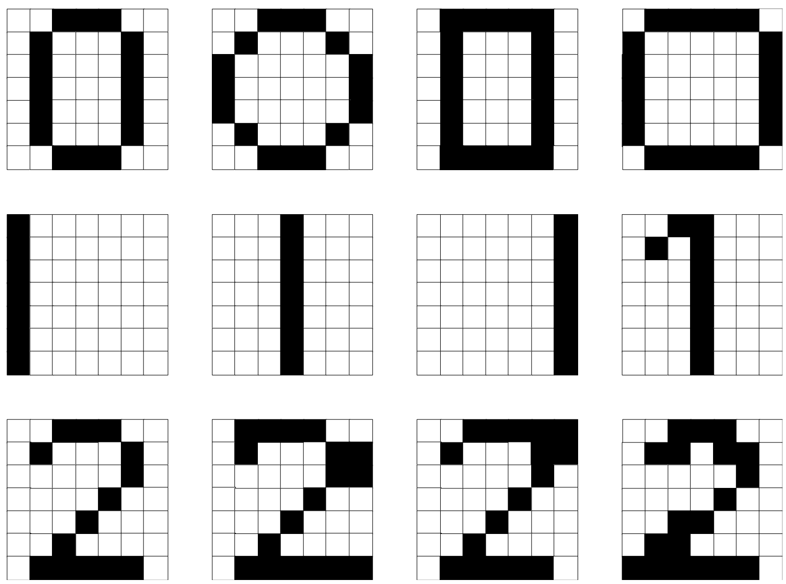


Figure 3.1: Symbols for numbers 0, 1 and 2.

Firstly, we indicate where the character line goes through each of the 49 sub-boxes (or better, when the line "covers" more than 50% of the area of the sub-box). If the numeral line passes through a sub-box, the corresponding value in the model vector is set to "1"; otherwise the corresponding value is left set to "0".

As mentioned before, model vectors may be prepared in two different ways. In the first case, the model vector is constructed of columns, where the columns are indexed from the left to the right. Such a model vector has 49 input variables (a 7 x 7 mesh) and three output variables (three different symbols), altogether 52 variables. In the second case each input variable represents a sum of the black pixels in the directions of both main diagonals of the box. Determined in such a way, the model vector has 26 input variables (13 sums in one direction + 13 sums in the other direction) and three output variables (again three different symbols!), all together 29 variables. A simple example from Figure 3.2 shows both ways of coding model vectors.

$\{\mathbf{mv}\}^T = \{1, 0, 0, 1, 1, 1, 0, 0, 0\}^T$ $\qquad$ $\{\mathbf{mv}\}^T = \{1, 1, 1, 1, 0, 0, 1, 2, 1, 0\}^T$

Figure 3.2: Two ways of coding model vectors in case of recognizing characters.

Preparation of the model vector in the second case shows the possibility of a simple reduction of the amount of necessary information. The number of input variables decreases from 49 to 29.

NOTE: There are two data bases. The first data base EXAMPL3A.CSV contains model vectors, which have 52 variables (the first way of coding model vectors). The second one EXAMPL3B.CSV contains model vectors, where the symbols are described with 29 variables (the second way of coding model vectors). The whole phenomenon (recognition between three different symbols) is described with 12 model vectors. Results in both cases are obtained using two files, EXAMPL3A.PRD and EXAMPL3B.PRD.

3.3 Verification of the Model

The efficiency of both models was established at different values of penalty coefficient. Figure 3.3 shows test examples - these are four noisy numbers (one "1", two "2" and one "0", respectively).

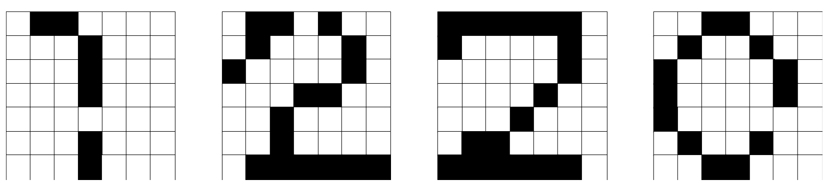


Figure 3.3: Noisy numbers 1, 2, 2 and 0.

aiNet will recognize all of numbers without difficulty. The correct predictions have extremely high values (certainties) in the case of low values for the penalty coefficient. Such a result is logical, and is expected, whilst the noisy numbers are still very similar to the given numbers in the data base.

High values for the penalty coefficient give lower certainties for the correct predictions, but still the highest in comparison to the other possible solutions. Even a very high value for penalty coefficient (e.g. 100.0, see Table 3.1) still gives the greatest value for the correct prediction in both models. The result shows an interesting fact, that the amount of necessary information may be reduced even more to achieve applicable results in the given example. This can have a practical and useful meaning in cases where we deal with a huge number of symbols and large vectors, in order to describe them.

Table 3.1: Results for both models at different values for penalty coefficient.

penalty coefficient	model 1				model 2			
	No. 1	No. 2	No. 2	No. 0	No. 1	No. 2	No. 2	No. 0
1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95
5.0	1.00	1.00	1.00	0.96	0.99	0.99	0.95	0.66
5.0	1.00	1.00	1.00	0.82	0.92	0.73	0.82	0.55
50.0	0.56	0.53	0.54	0.37	0.40	0.37	0.39	0.36
100.0	0.44	0.43	0.44	0.35	0.37	0.35	0.36	0.35

3.4 Comments

This example is of a more illustrative nature. It shows the simplest variants only; how to prepare the model vectors and the influence penalty coefficient to final a solution. We can say without qualification, that aiNet may be used for solving real problems in the area of pattern recognition, as in the recognition of either printed or hand-written characters. It may also be used for other problems in sensor processing, as data compression and noise removal from time-series signals.

Example 4

4. ECONOMY: Regional Analysis - Regional Development Model

4.1 General

In almost all countries there is a trend towards a decentralization of decision-making in politics as well as in economy. An economic and a political optimum may be imposed for a whole country but this optimum collapses with the introduction of regional optimums. All of those optimums should be taken into account but they can, however, cause a competition and disharmony among other regional interests and even disagreement with the state. It is therefore evident, that the question of (under)development of a region plays a very important role in the development policy of a country.

One of the initial problems in a regional analysis is a regionalization, i.e. dividing a particular area (country) into regions. Economists have no unified criterion - some do not define a region at all, some define its specific characteristics and some provide a concrete definition. We will not deal with this problem here, we shall suppose that it's solution is already known. It is our purpose to develop a model on the basis of which (and with the help of the most important variables) we will be able to distinguish between a developed and underdeveloped region and also the level of it's (under)development.

4.2 Modeling

When dealing with a problem mathematically, economists are usually unable to define exactly why they came to a certain conclusion. On the other hand, they can easily provide conclusions in some cases (e.g. a region is undeveloped, if its GNP is lower than 1). Therefore, we first drew up a questionnaire in view of five most important indicators of regional development. The questionnaire has been filled out by several experts. On the basis of the completed questionnaires, we prepared a data base, which when abridged, serves as a test example. There are more techniques to be implemented in the preparation of the data base and the final model. They will be included future aiNet examples.

The most important indicators signifying the level of regional development (*LRD*) are[+] :

- general national product per capita - *GNP*
- a public sector investment share in a state investment (%) [*I_invest*]

[+] The question of indicators presents a special problem and is not dealt with here specifically. The point is, we want to show how successful a method can be, however, the number and importance of indicators plays no vital role in it.

- population growth index (number of new-born babies in one year / number of deaths in one year * 100) [*I_pop*]
- a portion of students at colleges and universities in the population, expressed by indices (number of students at colleges and universities / number of all inhabitants * 100) [*I_stud*]
- a portion of employees compared in the population, expressed by indices (number of employees/ number of all inhabitants * 100) [*I_employe*]

A model vector may in general be expressed as follows:

$$\{\mathbf{mv}\}^T = \{GNP, I\_invest, I\_pop, I\_stud, I\_employe; LRD\}^T$$

NOTE: There is only one data base for the above example - EXAMPLE4.CSV. The phenomenon has been described by 243 model vectors, each of them is presented by six of the above variables.

4.3 Verification of the Model

In the model of regional development we take an interest in the development compared with other regions, therefore, the magnitude of the penalty coefficient is not importance. We should only ensure that the solution is smooth enough, as is generally the case in the nature. In using the model we suggest a penalty coefficient between 0.15 and 0.25. We can even choose a higher value but to the extent that the model will still not give constant values as a result.

4.4 Comments

The results obtained by the model may be regarded in two ways:

- we simply want to know if a region is developed (LRD > 5.5.) or underdeveloped (LRD < 5.5.) and /or
- we are interested in the LRD compared with the other regions. The LRD is important in case of different measures imposed by a state to encourage development in those regions

This model has some advantages over the usual statistical methods or over diagrams with quite simplified criteria. Let us mention the most important ones:

- we can take into account as many indicators as we know and can sufficiently evaluate
- the development can be observed by different combinations of indicators
- the model may be adjusted to new experience, new indicators may be added or the old ones which proved inadequate excluded
- reversed relations may be obtained easily : e.g. we can locate areas of investment needed to make a region a developed one (or leave it undeveloped) if other indicators are known (fixed).

Figure 4.1 shows in which way the LRD depends on two indicators: the *GNP* and investments. It should be mentioned that the helping tool, which enables the depiction of Figure 4.1 will be available in the next version of aiNet.

It is evident that the model should offer the known trends: the greater the *GNP* the higher the regional development, the greater the share of investments, the higher the regional development. It is also obvious that the *GNP* is a much more important indicator than investments. Correspondence of the results with the known quality relations only proves the usefulness of the model, which, on the other hand, enables the quantity evaluation of regional development.

Figure 4.1: Dependence of regional development on two most important indicators: GNP (horizontal) and investments (vertical)

Example 5

5. CIVIL ENGINEERING: Diagnosis of Damage of Prestressed Concrete Piles During Driving

5.1 General

In driving prestressed concrete piles (PCP), cracking and spalling may be encountered. The damage to such piles may be classified into three major types:

- spalling of concrete at the head of the pile due to high compressive stress,
- spalling of concrete at the point of the pile due to hard driving resistance at that point, and
- transverse cracking or breaking of the pile due to the combination of torsion and reflected tensile stress.

The damage depends on many factors, such as subsoil conditions, the features of the pile composition and inadequate or improper methods and equipment. While damage results in time delays, injuries and cost overruns, diagnosis of the causes of damage is very important.

Visual damages such as cracks and concrete spalls at different locations, together with the known characteristics of PCP indicate the causes of those damages. The judgment of causes is strongly based on human experts' knowledge. Usually, a human being needs years of study or practice to achieve the status of an expert.

Human expert solutions involve two problems: dependency on the expert and possible human subjectivity. The best way to avoid these problems is to formalize the expert knowledge from many sources and represent it in a mathematical form. In formalizing the expert knowledge, one encounters some major problems, including knowledge representation, reasoning and acquisition. Knowledge acquisition represents the most difficult task in this process. Usually, human experts cannot explain, how they reason and why they obtain such conclusions. Procedural solution can therefore not be used. The use of artificial neural networks seems to be a good alternative solution. The expert knowledge may be represented in the form of samples (rules) acquired from the human experts. Such samples consist of two parts: the first part represents features - descriptions of visual damages and other known characteristics of PCP during driving, the second part represents the causes of the damage.

5.2 Modeling

The knowledge base, as named the database in this case, was obtained directly from a paper published by Yeh et all [[5]]. 18 different visual damages and some known characteristics of PCP during driving were considered as follows:

[1] spalling of concrete occurred at the head of the pile
[2] spalling of concrete occurred below the head of the pile
[3] cracking occurred on one side of the pile
[4] cracking occurred around the pile
[5] cracking occurred at pile slit
[6] direction of cracking is longitudinal
[7] direction of cracking is transverse
[8] the slenderness of the pile is too high
[9] pile has splices
[10] driven times of cushion are many
[11] cushion is broken after pile-head breaking
[12] cushion is crushed after pile-head breaking is broken
[13] spiral reinforcement after pile-head breaking is broken
[14] ratio of driving formula resistance to design resistance is exceeded
[15] penetration of pile driving has ever increased suddenly
[16] penetration of pile driving has ever decreased suddenly
[17] pile hole after driving is full with soil and water
[18] pile hole after driving is full with concrete spall

Many different causes are possible during driving PCP. The most probable causes, which the model can propose as a solution, are as follows:

[1] uniformity of cushion material is poor
[2] ductility of cushion material is poor
[3] cushion is overused
[4] energy of hammer is to large
[5] driving in very hard rock or soil
[6] driving in very soft soil or cave
[7] concrete honeycomb
[8] strength of concrete is insufficient
[9] strength of pile splices is insufficient
[10] strength of pile tip is insufficient
[11] strength of prestress in insufficient
[12] strength of spiral reinforcement is insufficient

A uniform type of the variable was used to describe the phenomenon: each visual damage or known PCP characteristic is labeled as "0" or as "1"; the same convention is used in labeling possible causes. "1" means certainty of visual damage, certainty of known PCP characteristics or certainty of possible cause, and vice versa for "0". For example, if cracking occurred on one side of the pile, the

---

[5] Yeh, Y., Kuo, Y. & Hsu, D**., Building KBES for diagnosis PC pile with artificial neural network**, Journal of Computing in Civil Engineering, Vol. 7, No. 1, pp. 71 - 93, 1993.

third component of the model vector has value a ‘1’. Finally, each sample in the knowledge base is presented as a model vector, which has components of either zeros or ones.

Model vector

$$\{\mathbf{mv}\}^T = (0,1,1,0,0,1,0,1,0,0,1,0,0,0,0,0,0,0;\ 1,0,0,0,0,0,0,0,0,0,0,0)^T$$

represents the situation, *where spalling of concrete occurred below the head of the pile*, *cracking occurred on one side of the pile*, *direction of cracking is longitudinal, the slenderness of the pile is too high*, and *the cushion is broken after pile-head breaking*. A possible cause is that *the uniformity of cushion material is poor*.

NOTE: Database or knowledge base in this case, is written in the EXAMPLE5.CSV file. The phenomenon is completely described by 240 model vectors. Each model vector has 30 variables, where 18 of them are input variables and the rest (12) are output variables.

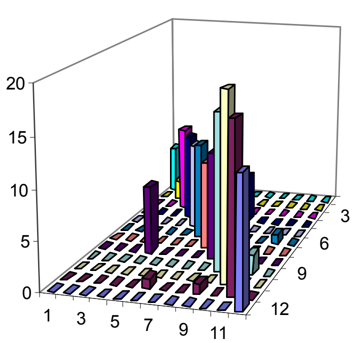
5.3 Verification of the Model

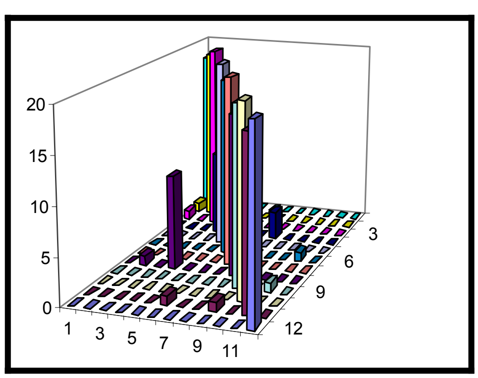
Usually, available data bases for certain problems do not contain representative samples. Non-representative samples do not "cover" all possible cases - all possible combinations of input variables. An optimal solution, based on such data is therefore not the actual optimal solution, but the optimal solution from an instantaneously available data base. The penalty coefficient was determined through iteration process. The optimal prediction error was defined as the point where the number of incorrect predictions is minimized.

The analysis of the problem shows that very small values of penalty coefficient give always only one cause for each sample from the data base (which is a reasonable result). The value of output variable ascribed to this cause is almost equal to one. On the other hand, higher values of penalty coefficient give values greater than zero for almost all output variables - the possible causes. These values may be interpreted as certainties for each cause.

An "exact" solution is defined as one, where the correct cause has the highest value among all possible causes. It was determined that the best value of penalty coefficient should be between 0.20 in 0.50.

The results of analysis are shown first as histograms of prediction errors, which are combinations of two presentations. A ground-plan of histograms shows correlation (or comparison) between predicted and actual results, while axonometric view shows the number of correct predictions for all possible causes (see Figure 5.1). Dependency of the optimal solution on the penalty coefficient is shown in Figure 5.2.





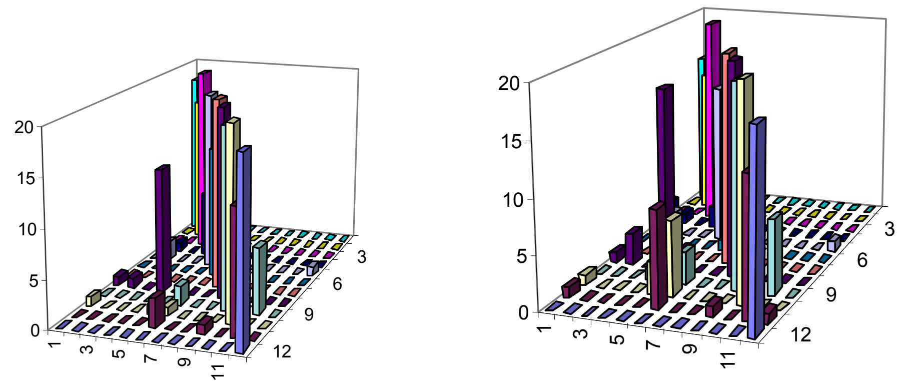


Figure 5.1. Histograms of prediction errors at different values for penalty coefficient.

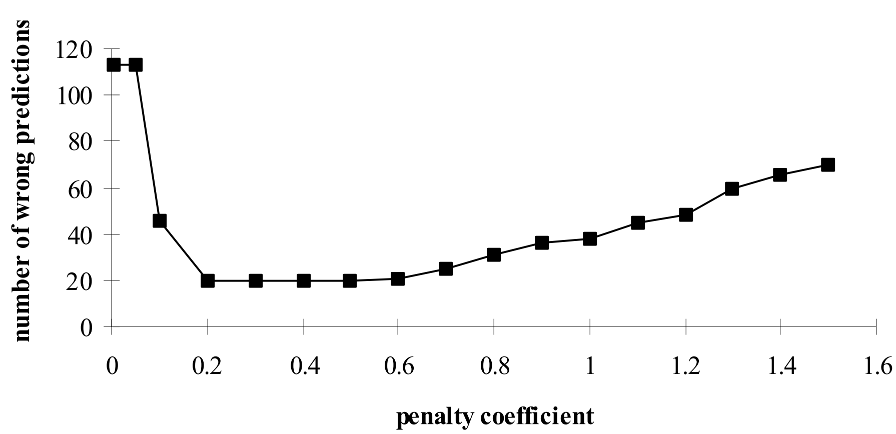


Figure 5.2. Dependency of the optimal solution on the penalty coefficient.

Figure 5.2 shows the same number of exact predictions for penalty coefficient values between 0.2 and 0.5. While the prediction is better, there is less disagreement between predicted and actual results and the value (certainty) is higher for actual cause in case of wrong prediction, the optimal solution is obtained at penalty coefficient 0.5 (see Figure 5.1./b/). We propose this when aiNet is used in practice for diagnosis of damage of PCP during driving.

NOTE: aiNet uses two different types of normalization: *regular* and *statistical*. In the case of regular normalization, the model vectors are automatically transformed linearly from real world to abstract hyperspace with dimensions (-1, 1) for each variable. Nevertheless, the results of this example are obtained on the basis of transformation to abstract hyperspace with dimensions (0, 1) for each variable. The use of different hyperspaces does not influence the solution significantly, only the optimal value of penalty coefficient is different.

5.4 Comments

The use of neural networks and/or neural network-like intelligent systems shows practical experiences and very good results in the field of civil engineering, too. aiNet in this case was used for diagnosis of damage of PCP during driving. The example represents a special case of qualitative reasoning, which differ from quantitative reasoning predominantly in the way of coding the data base (or knowledge base). We can conclude, that technical applications demand an appropriate tolerance level, which assure the optimality of the solution. Generally, we must not be too precise (large number of digits in results of analysis). It should be noted that precision of the model and optimal solution depend not only on penalty coefficient, but also on the size and quality of the data base. In connection with human logical reasoning, this means the result of reasoning is more certain with more and better information about the phenomenon.

Example 6

6. MEDICINE: Diagnosis in Case of Back Pain

6.1 General

Most of the people have back pain from time to time. Doctors do not usually try to determine the nature of the complaint from the observation of symptoms, while pain usually disappears after short repose. Doctors indicate such pain as an unspecified back pain. This complaint is the main reason for lost working days throughout the world [[6]]. People, who carry or lift up heavy objects are prone to a back pain. Not only they, but also the others who sit a lot in the same position or those who have to be crooked in an unnatural position, may have problems with their back. In any case, back pain may have many causes. The correct diagnosis of back pain (as is the case with all symptoms) is of importance to the health of the patient.

The example, which shows diagnosis of a recent back pain, serves as an illustration for the use of aiNet. In such cases, one usually decides to make a use of expert system; we will show how to use aiNet as representative of neural networks in this case.

It should be mentioned that expert systems have one advantage compared to neural networks, which might be of importance: a relatively simple procedure allows explanation of all steps in the reasoning process (consultation!). Neural networks have, on the other hand, capability of generalization, which might also be of importance in most cases of diagnosis: solutions may, for instance, be obtained for the samples which are not explicitly codified in the knowledge base.

6.2 Modeling

Most of natural science problems, especially problems which can be measured, have simple relations between their quantities. These relations are of the same kind and are described mathematically by operators (only "***and***", only "***or***", only "***and/or***"). Such operators are captured automatically in the knowledge base. On the other hand, a diagnosis operates with quantities which are connected with one another by different kinds of operators. Coding the data base (preparation of the model vectors) has become a little bit more complicated than in other cases. Let us try to analyze the simple part of the decision tree, which is shown in Figure 6.1.

---

[6] Smith, T., **Complete Family Health Encyclopedia**, Dorling Kindersley Limited, London, 1990.

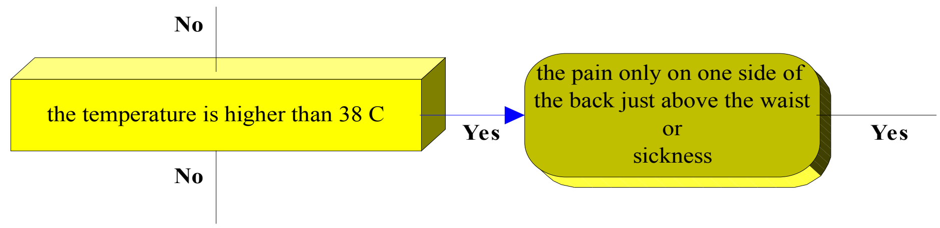


Figure 6.1: Part of the decision tree in case of diagnosis of the back pain.

Take a simple example, using two model vectors to describe the problem, using the convention that the quantity has value "1", if it is real, and "0" otherwise.

According to the above convention, the case from Figure 6.1 can be written in form of three model vectors as follows:

{..., 1, 1,0, ...},

{..., 1, 0,1, ...}, and

{..., 0, ...}.

The first two records (model vectors) convey:

..., *IF*

the patient has a temperature higher than 38ºC and he/she feels the pain just on one side of the back

or

he/she feels feebleness,

THEN ...

The third record (model vector) conveys:

..., *IF*

the patient has a temperature less than 38ºC,

THEN ...

A simple diagnosisfor back pain is shown in Figure 6.2. The decision tree from this figure gives the basic information for the preparation of knowledge base. Input variables (symptoms of the problem) of the model vectors are as follows:

1.1 the pain after lifting up a heavy burden, after a cough
and/or
1.2 after an exhausting physical exercise
2 the temperature is higher than 38ºC
3.1 the patient is over 60 years old
and/or
3.2 the patient spent several weeks in bed or in a wheel chair
4 the patient is over 45 years old

5	the pain is worse in the morning when getting up
6.1	the patient is hindered by the pain when moving or
6.2	has a stitch in one leg
7	the pain is localized mainly in the back and does not spread anywhere else
8.1	the pain is only on one side of the back just above the waist and
8.2	sickness
9	the pain is much worse on one side of the backbone
10	the pain occurs mostly in the neck or higher in the back between shoulder blades

Output variables of the model vectors - diagnosis have the following meaning:

1	SCIATICA, caused by pressure on a root of a kidney nerve; consultation with the doctor
2	a possible LUMBAGO, resulting from various exertions
3	a possible infection of kidneys - PYELONEPHRITIS, or the pain might be accompanying general virus inflammation; consultation with the doctor
4	back pains (maybe severe), resulting from virus infections, e.g. INFLUENZA; consultation with the doctor
5	an immediate consultation with the doctor, a possible bone injury
6	a possible ARTHRITIS in the neck vertebrae; a consultation with the doctor
7	a possible OSTEOARTHRITIS in a lower breast, renal or back part of the backbone; a consultation with the doctor
8	the patient's bed might not give him enough support, a chronic inflammation of joints is possible, as a result of a SPONDYLITIS; a consultation with the doctor
9	a consultation with the doctor

NOTE: The data base also named the knowledge base in the given example, is stored in EXAMPLE6.CSV. The entire phenomenon - diagnosis of recent back pain is represented by 18 model vectors. Each model vector has 23 variables, where 14 of them are input variables, and 9 are output variables. The results shown in the next subsection and are obtained on the basis of (prediction) vectors, stored in EXAMPLE6.PRD.

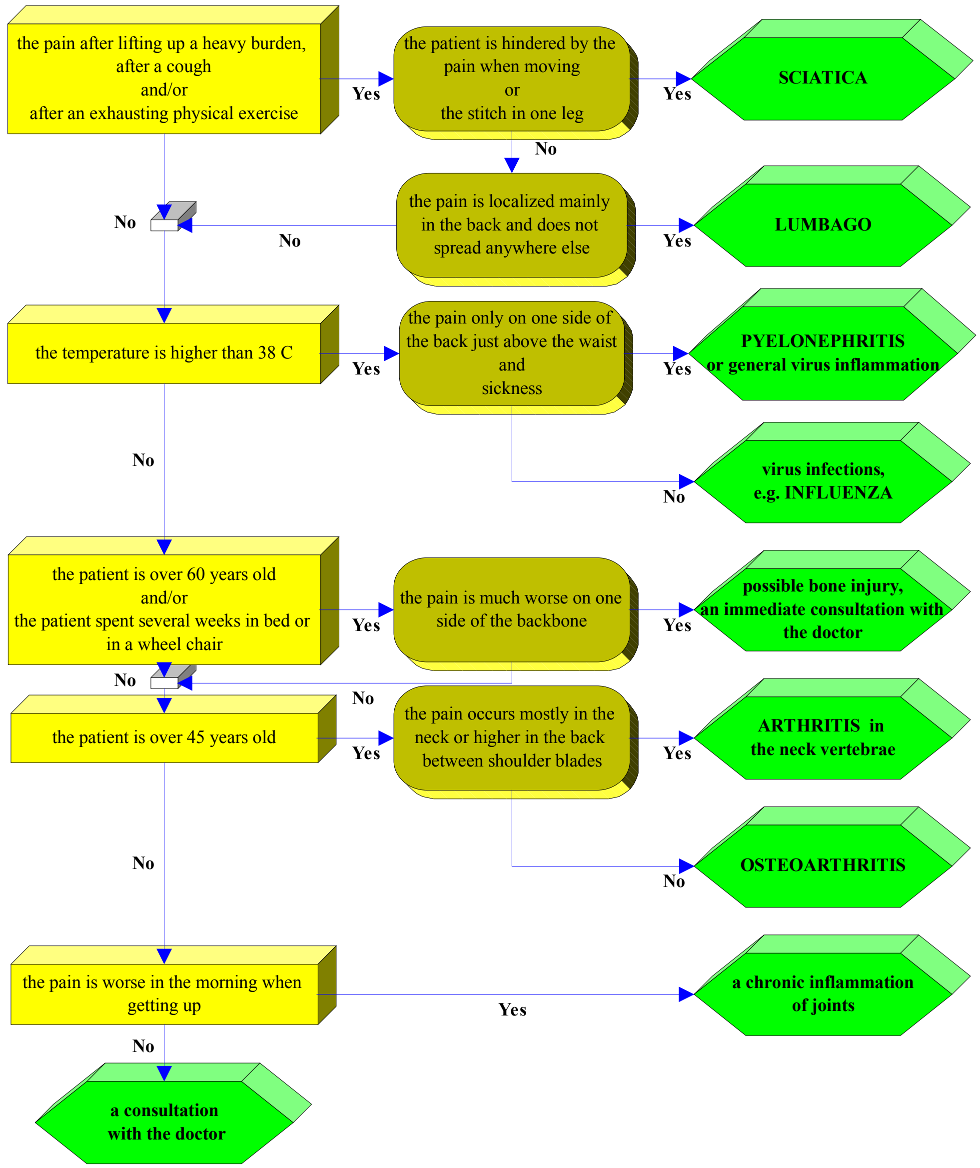


Figure 6.2: Decision tree in case of back pain diagnosis.

6.3 Verification of the Model

The model is not verified in the true sense of the word. Moreover, it is controlled just for a single value (0.5) of penalty coefficient. Different combinations of symptoms are used as test examples. The first two are the first two samples from the knowledge base, while the other two are hypothetical samples as follows:

- pain appears after a patient has lifted up heavy furniture, after a cough,
  and
  after gymnastics
- the patient's temperature is higher than 38ºC,
- the patient is over 60 years old

- the patient's temperature is higher than 38ºC,
- the patient is over 60 years old
  and
  he has spent a few weeks in bed recently

The results may be interpreted as certainties for each diagnosis, as we have done this in the case of diagnosis PCP during driving (see Example 5). The test results show that the certainties are practically equal to one for SCIATICA in the first two samples (which is the correct diagnosis); all other diagnoses have values practically equal to zero. As we can see, the model is able to predict the learning cases. (Users can check, if the model also predicts other samples from the knowledge base.)

The third hypothetical sample has about 50% certainty for SCIATICA*,* about 25% for LUMBAGO and about 25% for virus infection (e.g. INFLUENZ*A*). The fourth hypothetical sample has about 50% certainty for virus infection (e.g. INFLUENZA) and about 50% for OSTEOPOROSIS

6.4 Comments

Only doctors can assess the validation of the aiNet model in the given example, so we leave the final assessment to the experts. It should be mentioned again, that the given example is of illustrative nature. Practical use of similar and/or of more sophisticated models demands a user friendly interface for pre-processing the knowledge base and post-processing the predicted results. The data from decision trees will be automatically transformed in and out of binary form, which aiNet then uses for prediction.

The authors strongly believe that aiNet is a powerful tools which may help to prepare highly sophisticated models in medical diagnosis. Such models can help doctors in the practice, especially in cases of more complex diseases. The technique is appropriate for coding and presentation of the knowledge, which has not been explicitly recorded. There are also some other possible applications, for example recognition of genetic codes, etc.

References

1. Carpenter, W., C. and Barthelemy, J., F., **Common misconceptions about neural networks as approximators**, Proceeding of the 3rd International Conference on the Application of Artificial Intelligence to Civil and Structural Engineering, in Neural Networks and Combinatorial Optimization in Civil and Structural Engineering (Topping, B. H. V., editor), pp. 11-18, 1993.

2. Hecht-Nielsen, R., **Neurocomputing**, Addison-Wesley Publishing Company, Inc., 1990.

3. Guyon, L. et al, **Comparing different neural network architectures for classifying handwriting digits**, Proc. of the Int. Joint Conf. on Neural Networks, **II**, pp. 127-132, IEEE Press, New York, June 1989.

4. Hecht-Nielsen, R., **Nearest matched filter classification of spatiotemporal patterns**, Applied Optics, **26** (10), pp. 1892-1899, 1987.

5. Yeh, Y., Kuo, Y. & Hsu, D**., Building KBES for diagnosis PC pile with artificial neural network**, Journal of Computing in Civil Engineering, Vol. 7, No. 1, pp. 71 - 93, 1993.

6. Smith, T., **Complete Family Health Encyclopedia**, Dorling Kindersley Limited, London, 1990.