An Application of Unsupervised Neural Networks in General Insurance: the Determination of Tariff Classes

Renato Pelessoni – Liviana Picech

Dipartimento di Matematica Applicata
alle Scienze Economiche Statistiche ed Attuariali
"Bruno de Finetti", Università degli Studi di Trieste

Abstract: *This paper concerns the application of some type of unsupervised neural networks to the specific actuarial problem of collecting the values of one tariff variable into tariff classes. We show how these techniques allow applying partitioning methods of cluster analysis to this specific problem and investigate the possibility of taking advantage of a topological property of Kohonen self-organising maps in order to build tariff classes containing contiguous values of the tariff variable. A numerical application shows the procedure.*

Keywords: *Tariff classes, unsupervised neural networks, competitive learning, Kohonen self-organising maps.*

1. Introduction

In recent years Neural Networks (NN) have been having a wide spread of applications in many different fields. In the actuarial practice, Lowe and Pryor (1996) have reported on the application of supervised NN in underwriting, since this type of NN is specifically designed to deal with models representing a set of information from which some sort of predictions are derived.

This paper is concerned with unsupervised neural networks. As the supervised NN are connected to statistical models for predictions, the unsupervised NN are connected to cluster analysis techniques. Giulini, Pelessoni and Picech (1997) have applied some types of unsupervised NN to collect the values (basic classes) of one tariff variable into tariff classes; it has been showed how these techniques allow implementing partitioning methods of cluster analysis, taking also account of

---

Corresponding author: Renato Pelessoni, Università degli Studi di Trieste, Piazzale Europa n.1, 34127 Trieste – Italy – Tel.:+39-040-6767062 – Fax:+39-040-54209 - email: renatop@econ.univ.trieste.it.
This research work was partially supported by M.U.R.S.T. (Progetto di ricerca scientifica di rilevante interesse nazionale – 1998 – prot. 9813564672_006: "Models for the management of financial, insurance and operations risks").

different exposure of the basic classes and producing appreciable results in comparison with the traditional actuarial methods.

In this paper, we still deal with the problem of collecting the basic classes of one tariff variable in clusters. In particular we investigate the possibility of taking advantage of a topological property of Kohonen Self-Organising Maps in order to build tariff classes containing contiguous values of the tariff variables.

An outline of the paper is the following.

In Section 2 the problem of determining tariff classes in rate making is delineated.

In Section 3 we briefly describe two neural network algorithms frequently used in clustering problems: Simple Competitive Learning and Kohonen Self-Organising Map.

Sections 4 is devoted to an application of the algorithms described in Section 3 to collect in clusters the values of the tariff variable “age of the insured” in a motor vehicle insurance portfolio.

In Section 5 the topology preservation property of the Self-Organising Maps is exploited to collect the basic classes described by the age of the insured in a motor insurance portfolio in clusters formed by contiguous values of the tariff variable. A heuristic procedure is introduced and it is described throughout a numerical example.

2. The determination of tariff classes in general insurance rate making

In general insurance, the premiums for insurance covers are often determined by means of tariffs. A tariff defines the insurance premium as a function of some observable variables describing the risk. For instance, in motor vehicle liability insurance the premium can be determined as a function of the power of the vehicle, of the geographical area where the insured lives, of the age of the insured, etc. These observable variables are called tariff variables. In rate making, we use statistical methods, mathematical algorithms and practical reasoning to build the tariff structure; in this way the statistical information on the claim experience is combined with the observation on the tariff variables. The premium for a new risk can then be determined as a function of the observed values of the tariff variables.

Many tariff methods require the values of the tariff variables to be collected in classes and, even if this is not necessary, commercial reasons often suggest making use of a low number of tariff classes. For instance, the values of the tariff variable "age of the insured" in a motor vehicle insurance tariff can be grouped into the age classes 18-26, 27-35, 36-43, 44-60 and over 60. Obviously, a question arises: which values of the tariff variable should be grouped together and which not, and also how many classes should be formed.

In the actuarial literature some cluster analysis techniques have been applied to collect the values of one tariff variable (the so called basic classes) in clusters, that is to say in homogenous groups, according to one or more characteristic variables describing the claim experience (see van Eeghen, Greup and Nijssen (1983) for a review). Turning back to the tariff variable "age of the insured", the claim experience can be described, for instance, by the claim frequency and ages with "quite similar" claim frequency have to be allocated in the same cluster.

The basic classes can then be seen as objects that have to be joined together, building the tariff classes, in accordance with the observed values of the characteristic variable; from this point of view the problem of determining the tariff classes can be seen as a clustering problem. However, the observed values of the characteristic variable in each basic class arise from observations on risks with different exposures (e.g. the number of observed policy-years in a motor vehicle insurance). Therefore, these values are not immediately comparable by means of the similarity or dissimilarity measures considered in traditional clustering procedures. For this reason, in the actuarial literature some clustering techniques have been implemented in order to take account of the exposures of the basic classes as well.

For instance, the method proposed by Loimaranta, Jacobsson and Lonka (1980) is a non-hierarchical method of mixtures in which it is assumed that the basic classes belong to a fixed number K of clusters and that the characteristic variables are independent random variables with probability distribution a mixture of K distributions, one for each cluster. In this method, the different exposures are dealt by means of the definition of these probability distributions. After the parameters of the distributions have been estimated by the maximum likelihood method, the posterior probabilities for each basic class to belong to the different clusters can be estimated and the allocation of the basic classes to the clusters is done according to these posterior probability distributions.

Another method has been proposed by Dickmann (1978). It is a hierarchical agglomerative clustering method in which, at the beginning, each basic class is viewed as a group containing one object and, at each stage, the merging of two groups is done if it minimises the increase of the total within-cluster variance. The procedure is repeated until all basic classes are located in one cluster.

For a short description of the algorithm, let us consider a single stage with the basic classes joined together in *K* clusters. Let

S_j	be the set of basic classes located in cluster *j*;
x_i	be the observation of the characteristic variable with respect to the *i*-th basic class;
t_i	be the value which reflects the exposure of the *i*-th basic class (e.g. the number of observed policy-years);
$n_j = \sum_{x_i \in S_j} t_i$	be the total exposure in cluster *j*.

Define the total within-cluster variance with *K* clusters as:

$$(2.1) \qquad {}^{(K)}\sigma_W^2 = \sum_{j=1}^{K} \sum_{x_i \in S_j} (x_i - \bar{x}_j)^2 \frac{t_i}{n}$$

where

$$\bar{x}_j = \sum_{x_i \in S_j} x_i \frac{t_i}{n_j} \quad \text{and} \quad n = \sum_{j=1}^{K} n_j \ .$$

It is important to note how the definition of within-cluster variance allows taking account of the different exposures of the basic classes. We pass from *K* to *K-1* clusters by merging two of the existing clusters so that the increase of the within-clusters variance ${}^{(K-1)}\sigma_W^2 - {}^{(K)}\sigma_W^2$ is minimum.

Another important class of cluster analysis techniques is known as partitioning methods (among which the well-known k-means algorithms). In these methods the number of the clusters *K* is fixed in advance or, in some variants, determined through the procedure. Moreover, unlike the hierarchical techniques, they allow the relocation of the objects. In this way, bad initial partitions can be improved later. Most of these techniques consist of two distinct procedures:

– the determination of an initial allocation of the objects into the clusters;

– the relocation of some or all of the objects in the clusters.

An essential feature of these methods is the calculation of the centroids of the clusters. Many clustering algorithms have been proposed; among them those proposed by Forgy, by MacQueen and a variant of the latter method (see Anderberg (1973)) are reported in Giulini, Pelessoni and Picech (1997).

Loimaranta *et al*. (1980) stated that, in their opinion, as far as the determination of tariff classes is concerned, a method that searches for the optimal partition could be preferred to a hierarchical clustering technique. However, partitioning methods face the difficulty of considering the exposures of the basic classes. In Section 2, we will see how, in a NN framework, some partitioning algorithms can be implemented in a more flexible environment, allowing the exposures to be managed as well.

3. Self-Organising MapS and Simple Competitive Learning

The Self-Organising Map (SOM) algorithm was introduced by Kohonen (1984). The algorithm is implemented by means of a network (the SOM) whose vertices (units or neurons) are disposed in a lattice (generally a one- or two-dimensional array of units).

Denote by S a set of N vectors of $\Re^n$ (input space). Data are randomly chosen from S and we will denote with $x(t)$ the input vector drawn at time t ($t \geq 0$). An n-dimensional weight vector m_j is associated to each unit of the network, so that the state of the network at time t will be represented by $m(t) = (m_1(t), \ldots, m_K(t))$, where K is the number of units and $m_j(t)$ denotes the weight vector of unit j at time t.

If we present a vector of data $x \in S$ to the network, it can be compared with all the weight vectors. We call winner unit, $c = c(x)$, the unit satisfying the condition

$$d(x, m_c) \leq d(x, m_j) \ \ \forall j = 1,..., K$$

where d is a distance in $\Re$ (usually euclidean).

Let d' be a distance in the lattice and λ_t be a family of positive non-increasing real functions defined on $\Re^+$, with $\lambda_t(0) = 1 \; \forall t$.

The SOM algorithm (Kohonen (1995)) is carried out by adapting the weight vectors of the network by means of an unsupervised learning process. The algorithm consists of the following steps.

Self-Organising Map algorithm

<u>Step 1:</u> *put t =0 and initialise the vectors $m_j(0)$* $(j = 1, \ldots, K)$;

<u>Step 2:</u> *choose an input vector* $x(t) \in S$;

<u>Step 3:</u> *find the winner unit c=$c(x(t))$;*

<u>Step 4:</u> *update the weight vectors according to the rule*

$m_j(t+1) = m_j(t) + \alpha(t)\, \lambda_t(d'(c, j))\,(x(t) - m_j(t)) \quad (j = 1, \ldots, K)$ *with*

$\alpha(t) > 0$;

<u>Step 5:</u> *stop if the stopping rule is satisfied;*

otherwise replace t with $t+1$, go back to <u>Step 2</u> and repeat for the next input vector.

Several methods of initialisation of the weight vectors have been proposed in literature. A typical one is the so-called “random guess method”, in which the initial values are chosen randomly in the “right” domain, according to the values of the input vectors. Note also that a convenient stopping rule has to be fixed. The most common one consists in fixing in advance a “sufficiently” large number of iterations.

Usually, the term $\alpha(t)$ (learning rate factor) is a positive non-increasing function of t and $\alpha(0)$ is chosen not too far from 1 (typically 0.8). Also λ_t is a non-increasing function and, as a consequence, the weights of the units of the lattice close to the winner unit and those of the winner unit itself are changed significantly. On the other hand, weights of units placed further away from the winner unit are not updated appreciably.

At the end of the learning process the network can be used as a vector classifier.

Input vectors that make winner the same unit *j* are assigned to the same cluster S_j and the corresponding weight vector m_j can be chosen as "representative" of the cluster itself. Clearly, different runs of the algorithm can produce different results.

The choice of the functions λ_t is crucial for another important feature of the SOM algorithm, namely the property of topology preservation. Essentially, at the end of the learning process and with a right choice of the parameters of the algorithm, the neighbourhood relations are conserved, so that input vectors that are close in the input space are assigned to clusters represented by weight vectors of units which are close in the lattice.

A well-known expression for the function λ_t, widely used in the applications, is the gaussian function:

$$\lambda_t(r) = \exp\left(-\frac{r^2}{2\sigma^2(t)}\right)$$

where σ is a decreasing function and $\sigma(0)$ is large enough.

In the very specific case when

$$\lambda_t(r) = \begin{cases} 1 & if\ r = 0 \\ 0 & if\ r \neq 0 \end{cases}$$

that is if, at each time *t*, only the weight vector of the winner unit is updated, the so-called Simple Competitive Learning (SCL) algorithm is implemented. It is important to note that in this case the topology preservation property is no more in force. A thorough discussion on SCL can be found in the book by Hertz, Krogh and Palmer (1991), where also the strong relationship between SCL and k-means is pointed out.

In the SCL algorithm, the training is continuous, since the weights are updated after the presentation of each pattern (see Step 4 of the SOM algorithm). Nevertheless, there exists also a batch version of the same algorithm, proposed by Linde, Buzo and Gray (1980) and known as Linde-Buzo-Gray (LBG) algorithm of Vector Quantisation, where the weights are updated after all patterns have been presented.

Linde, Buzo and Gray proved that, if we denote by $p_i \ (i = 1,..., N)$ a probability distribution over the input space S and the input vectors are selected according to this probability distribution, the LBG algorithm converges to a local minimum of the quantity (average distortion)

$$(3.1) \qquad D = \sum_{j=1}^{K} \sum_{x_i \in S_j} d(x_i, m_j)^2 \, p_i$$

(see also Black (1992) and Luttrell (1990)). Note that, from an essentially practical point of view, continuous training is frequently preferred to batch training, because the random presentation order of the input vectors can help to avoid poor local minima (see Hassoun (1995) at page 168). Moreover, de Bodt *et al*. (1997, 1999) observed that often, in practical applications of SOMs, only the weight vector of the winner unit is updated in the final iterations of the algorithm and that the Kohonen algorithm can be considered an efficient initialisation procedure of SCL.

It must be noted that, despite the extensive use of SOMs, the mathematical theory of Kohonen's algorithm is so far unsatisfactory. A review on main mathematical results on the convergence of the algorithm and on the property of topology preservation has been reported by Cottrell, Fort and Pagès (1998). See also Bishop (1995) and Ripley (1996) for a wide investigation of the connections between neural networks and pattern recognition.

It is well-known that these unsupervised NN implement partitioning methods of cluster analysis. Comparing (2.1) and (3.1), Giulini, Pelessoni and Picech (1997) emphasised how a suitable definition of the probability distribution $p_i \ (i = 1,...,N)$ over the input space S of the basic classes allows implementing, by means of these algorithms, a partitioning method of cluster analysis which takes account of the different exposures of the basic classes. This can be done by defining the probability distribution as the relative exposure

$$(3.2) \qquad p_i = \frac{t_i}{\sum_{h=1}^{N} t_h} \qquad (i = 1,...,N)$$

where t_i is the exposure of the basic class *i*.

4. An example

In this Section we present the results obtained by applying the SOM and the SCL algorithms to the data in Table 4.1, where the claim frequencies in a motor vehicle insurance portfolio are reported (see also Figure 1, where the data are shown in a graph). In this example, our aim is to collect the ages of the policyholders (basic classes) in clusters, according to their claim frequencies. To perform the experiments, we used Matlab and the Neural Network Toolbox, even though the original programs have been substantially modified to implement our applications.

In the SCL procedure, as suggested by Kohonen (1995), an individual learning rate was assigned to each weight vector, by means of the recursive formula

$$\alpha_j(t+1) = \begin{cases} \dfrac{\alpha_j(t)}{1+\alpha_j(t)} & \text{if } j = c(t) \\ \alpha_j(t) & \text{if } j \neq c(t) \end{cases} \qquad (j = 1,\ldots,K)$$

where $\alpha_j(t)$ is the learning rate assigned to unit *j* at time *t* and *c*(*t*) is the winner unit at the same time. In this way, in every training cycle, only the learning rate corresponding to the winner unit is updated.

In the SOM experiments a network with a one-dimensional array of units and the euclidean distance was considered; the gaussian function was used in <u>Step 4</u> of the algorithm. Besides, α and σ were defined as suggested by Ritter and Schulten (see the book by Hassoun (1995) at page 114):

$$\alpha(t) = \alpha_0 \left(\frac{\alpha_{t_{\max}}}{\alpha_0} \right)^{\frac{t}{t_{\max}}} \qquad\qquad \sigma(t) = \sigma_0 \left(\frac{\sigma_{t_{\max}}}{\sigma_0} \right)^{\frac{t}{t_{\max}}}$$

where $t_{\max}$ is the maximum value for *t* (fixed in advance) and α_0, $\alpha_{t_{\max}}$, σ_0, $\sigma_{t_{\max}}$ are the fixed initial and final values of α and σ respectively ($\alpha_0 = 0.8$, $\alpha_{t_{\max}} = 0.01$, $\sigma_0 = 0.75$, $\sigma_{t_{\max}} = 0.25$ in our experiments).

In Table 4.3 are reported the subdivisions in clusters showing the lowest value of distortion, obtained by means of the SCL and the SOM algorithm for different numbers of units. To simplify the description of the clusters, the basic classes and the claim frequencies, ordered by the latter, are reported in Table 4.2.

Table 4.1: Policy-years (exposure) and relative and absolute claim frequencies in automobile insurance for different policyholder's ages.

Age	No. of Claims	Exposure	Claim Frequency	Age	No. of Claims	Exposure	Claim Frequency
18	23	91,01	0,252706	57	219	2073,21	0,105633
19	113	593,24	0,190480	58	187	1746,44	0,107075
20	263	1266,33	0,207686	59	174	1714,64	0,101479
21	306	1939,81	0,157748	60	168	1637,98	0,102565
22	376	2156,65	0,174345	61	132	1498,30	0,088100
23	362	2566,41	0,141053	62	146	1450,78	0,100636
24	391	2724,07	0,143535	63	133	1442,04	0,092230
25	365	2832,83	0,128847	64	111	1390,82	0,079809
26	384	2974,93	0,129078	65	122	1329,39	0,091771
27	339	3132,37	0,108225	66	107	1135,92	0,094197
28	343	3177,86	0,107934	67	98	1035,29	0,094660
29	334	3311,80	0,100851	68	89	990,80	0,089827
30	327	3431,88	0,095283	69	80	922,72	0,086700
31	307	3418,68	0,089801	70	82	838,55	0,097787
32	302	3317,40	0,091035	71	82	787,63	0,104110
33	280	3087,49	0,090688	72	64	690,74	0,092655
34	300	3168,45	0,094683	73	60	590,56	0,101599
35	241	3016,52	0,079893	74	57	539,82	0,105591
36	254	2968,79	0,085557	75	44	434,87	0,101179
37	244	2860,14	0,085311	76	24	234,06	0,102537
38	225	2794,10	0,080527	77	15	177,68	0,084421
39	235	2831,11	0,083006	78	15	150,75	0,099502
40	229	2727,02	0,083974	79	16	171,77	0,093149
41	227	2819,13	0,080521	80	15	160,90	0,093229
42	245	2772,79	0,088359	81	10	122,52	0,081620
43	207	2582,16	0,080166	82	14	89,25	0,156859
44	263	2605,99	0,100921	83	8	66,34	0,120589
45	291	2737,60	0,106297	84	6	57,26	0,104794
46	247	2660,50	0,092840	85	1	44,51	0,022465
47	272	2764,65	0,098385	86	5	21,96	0,227728
48	264	2656,79	0,099368	87	3	22,47	0,133523
49	309	2728,08	0,113267	88	0	12,52	0,000000
50	199	2099,79	0,094771	89	1	12,71	0,078647
51	211	2103,36	0,100316	90	1	11,26	0,088826
52	250	2171,34	0,115136	91	0	7,00	0,000000
53	222	2068,38	0,107331	92	1	4,84	0,206782
54	215	2056,44	0,104550	93	0	2,90	0,000000
55	230	2221,64	0,103527	94	0	6,86	0,000000
56	203	2183,59	0,092966	95	3	33,14	0,090528

(Data provided by an Italian Insurance Company)

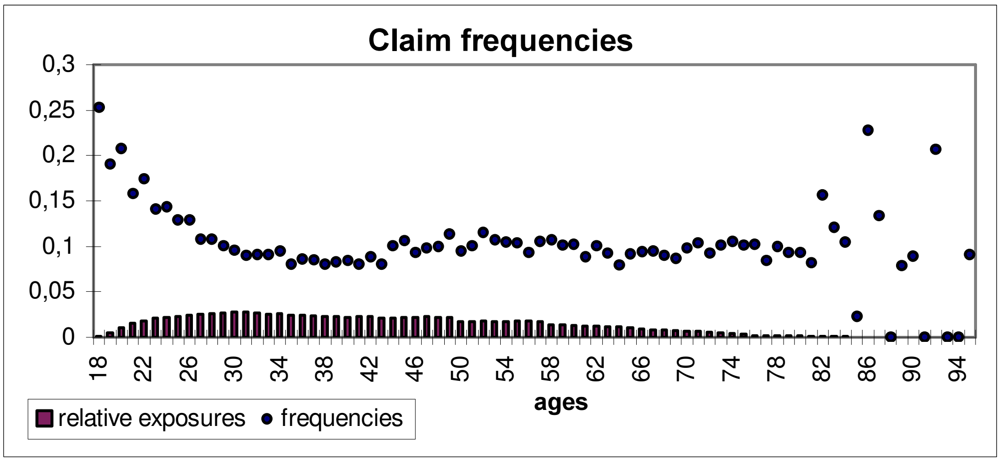

Figure 1

Table 4.2: Policyholder's ages and relative claim frequencies (ordered by claim frequencies).

Age	Claim Frequency	Age	Claim Frequency	Age	Claim Frequency	Age	Claim Frequency
18	0,252706	45	0,106297	30	0,095283	69	0,086700
86	0,227728	57	0,105633	50	0,094771	36	0,085557
20	0,207686	74	0,105591	34	0,094683	37	0,085311
92	0,206782	84	0,104794	67	0,094660	77	0,084421
19	0,190480	54	0,104550	66	0,094197	40	0,083974
22	0,174345	71	0,104110	80	0,093229	39	0,083006
21	0,157748	55	0,103527	79	0,093149	81	0,081620
82	0,156859	60	0,102565	56	0,092966	38	0,080527
24	0,143535	76	0,102537	46	0,092840	41	0,080521
23	0,141053	73	0,101599	72	0,092655	43	0,080166
87	0,133523	59	0,101479	63	0,092230	35	0,079893
26	0,129078	75	0,101179	65	0,091771	64	0,079809
25	0,128847	44	0,100921	32	0,091035	89	0,078647
83	0,120589	29	0,100851	33	0,090688	85	0,022465
52	0,115136	62	0,100636	95	0,090528	88	0,000000
49	0,113267	51	0,100316	68	0,089827	91	0,000000
27	0,108225	78	0,099502	31	0,089801	93	0,000000
28	0,107934	48	0,099368	90	0,088826	94	0,000000
53	0,107331	47	0,098385	42	0,088359		
58	0,107075	70	0,097787	61	0,088100		

The description of the subdivisions refers to the order in the data: e.g. (5 3 6 26 33 5) in Table 4.3 characterises the subdivision where the first cluster contains the first five elements in Table 4.2 (ages: 18, 86, 20, 92 and 19), the second cluster contains the next three elements (ages: 22, 21 and 82), etc.

We note that the distortions of the subdivisions obtained with the two algorithms (for the same number of clusters) are quite close and, except the subdivision into 6 clusters, the SCL clusters could appear preferable.

Table 4.3: Best subdivisions in clusters obtained by SCL and SOM.

No. of clusters	Clusters obtained by SCL	$D \times 10^{-5}$	Clusters obtained by SOM	$D \times 10^{-5}$
8	4 1 3 5 16 26 18 5	1.9423	6 4 4 9 17 17 9 12	2.0020
7	5 3 6 24 22 13 5	1.9692	6 4 4 9 17 20 18	2.1489
6	5 3 6 26 33 5	3.1042	6 7 10 17 20 18	2.9643

Analogously to the traditional partitioning methods of cluster analysis (e.g. k-means) we obtain different partitions of the basic classes depending on the stated number of clusters (units) K. Therefore, a problem arises: a criterion to decide how many clusters should be considered. Pelessoni and Picech (1997) applied, for this purpose, the method proposed by Schmitter and Straub (1975) to find the “best” subdivision of an insurance portfolio in tariff classes. They assumed the existence of a “natural subdivision” and derived two statistics to single out this subdivision, or possibly the “closest” one from a set of “admissible subdivisions” (the "admissible subdivisions" are a subset of all the subdivisions of the portfolio, which can be actually considered for practical and commercial reasons). We do not discuss here this criterion, but mention that, according to it, the subdivisions in 7 clusters seem to be the most preferable. The details of these subdivisions are reported in Table 4.4 and 4.5.

However, if we look at the resulting groups of basic classes (Tables 4.4 and 4.5), we realise that these subdivisions could be unsatisfactory for actual rate making purposes. In particular, the basic classes are not contiguously grouped and this fact is clearly unsatisfactory from a commercial point of view: e.g. (see Table 4.4) a 42-year-old insured should pay a premium different from that paid by a 41- or a 43-year-old. Moreover, in this procedure the information “age of insured” is not taken in any account. Observe also that the basic classes characterised by low exposures are anyhow classified according to their claim frequencies. For instance, the basic class "age 85" is classified in the 7th cluster, whereas the basic classes "age 84" and "age 86" are classified in the 4th and in the 1st cluster respectively, since these basic classes show very different claim experiences (see Table 4.1), even though their exposures are very low.

Table 4.4: Details on a subdivision in 7 clusters obtained by SCL.

Cluster	No. of elements	Policyholder's ages	Weights	Centroids	Exposures
1	5	18 19 20 86 92	0.2048	0.2048	1977.38
2	3	21 22 82	0.1662	0.1663	4185.70
3	6	23-26 83 87	0.1351	0.1353	11187.06
4	24	27-29 44 45 48 49 51-55 57-60 62 71 73-76 78 84	0.1050	0.1051	42389.66
5	22	30-34 42 46 47 50 56 61 63 65-68 70 72 79 80 90 95	0.0928	0.0927	38243.33
6	13	35-41 43 64 69 77 81 89	0.0824	0.0824	25225.41
7	5	85 88 91 93 94	0.0205	0.0136	73.78

Table 4.5: Details on a subdivision in 7 clusters obtained by SOM.

Cluster	No. of elements	Policyholder's ages	Weights	Centroids	Exposures
1	6	18 19 20 22 86 92	0.1880	0.1889	4134.02
2	4	21 23 24 82	0.1464	0.1466	7319.55
3	4	25 26 83 87	0.1288	0.1289	5896.57
4	9	27 28 45 49 52 53 57 58 74	0.1088	0.1088	20375.10
5	17	29 44 47 48 51 54 55 59 60 62 70 71 73 75 76 78 84	0.1012	0.1011	25617.76
6	20	30-34 42 46 50 56 61 63 65-68 72 79 80 90 95	0.0922	0.0921	34640.12
7	18	35-41 43 64 69 77 81 85 88 89 91 93 94	0.0827	0.0822	25299.19

This inconvenience could be avoided if, when grouping the basic classes, the information "age of the insured" would be considered as a substantial information and not only as a label attached to the basic classes just to identify them. An example of a procedure in which also the values labelling data are directly employed as source of information in clustering has been proposed by Jain and Farrokhnia (1991).

5. Clustering under a constraint of contiguous grouping

It would clearly be possible to take account of the actual value "age of the insured", in addition to the observed claim frequency, by applying the clustering techniques to the basic classes described by the couple of characteristic variables claim frequency and age of the insured. Therefore, since in this case the objects to be collected in clusters are two-dimensional vectors, a suitable distance in $\mathfrak{R}^2$ should be considered. The importance attached to the information "age" with respect to the observed claim frequency is determined by this distance. Nevertheless, this distance should also have the appreciable property of increasing this importance when the exposure (that is to say the number of observations) of the basic class "age of insured" is very low.

In this Section a different approach is presented. We develop a procedure in which the basic classes "age of the insured" are collected in classes under a sort of constraint of contiguous grouping. More precisely, successive applications of Kohonen SOMs are carried out and the property of topology preservation is exploited in order to induce the contiguous grouping in a natural way.

The procedure develops in two stages and it is illustrated by means of a numerical example in which the basic classes are actually described by the two characteristic variables "age of the insured" and "claim frequency".

I Stage

In the first stage two SOMs are trained: one concerns, as objects to be collected in clusters, the "ages of the insured" relative to the basic classes and the other the "claim frequencies". As a result we obtain classes of "ages" and classes of "claim frequencies". Thanks to the topology preservation property both the classes of ages and the classes of claim frequencies are ordered by age and by claim frequency respectively.

II Stage

In this stage we consider as initial basic classes the objects described by the two characteristic variables "age of the insured" and "claim frequency" and associate them the corresponding indexes of age class and of claim-frequency class obtained in the first stage. In this way, the new basic classes are described by two characteristic variables, the index of age class and the index of claim frequency class, and they are considered as the input space of another SOM. More precisely, the input space of this SOM is now a subset of $\Re^2$ whose elements are the couples (index of age class, index of claim frequency class) to which at least one initial basic class has been associated. Both the characteristic variables are now indexes of clusters resulting from the first stage and we consider the usual euclidean distance to compare the objects. Moreover, we assume as probability distribution over the input space the total amount of the relative exposures of the initial basic classes associated to each couple of indexes.

In the following, we report the results of an application of this procedure to the data in Table 4.1. We used the SOM algorithm by setting the parameters of the Ritter and Schulten formulas to the values $\alpha_0 = 1.5$, $\alpha_{t_{\max}} = 0.1$, $\sigma_0 = 2$, $\sigma_{t_{\max}} = 0.5$.

In the first stage one SOM is trained to collect the objects "age of the insured" into 20 clusters. The input space consists of the ages reported in Table 4.1 and the probability distribution has been defined by means of the relative exposures as in (3.2). Therefore, only the age values are considered, without any reference to the claim frequency. The resulting subdivision in clusters is reported in Table 5.1. Owing to the topology preservation property of the SOM algorithm, the indexes of the clusters are ordered according to the order in the input space. Moreover, the relative exposures of the resulting clusters are quite flat.

At the same stage the objects "claim frequencies" are collected into 9 clusters by another SOM. In this case the input space is the set of claim-frequency values from Table 4.1 and the probability distribution is defined again as the relative exposures as in (3.2). The resulting clusters are reported in Table 5.2 and also in this case the topology preservation property of the SOM algorithm makes their indexes ordered according to the claim-frequency values (cf. Table 4.2). In Table 5.2 are reported the labels “age of the insured”, to identify the objects collected in the same cluster. In the Second Stage one SOM is trained to collect objects described by the couple of indexes (age index, claim-frequency index) into 7 clusters, which, in our example, give an interesting result. The elements of the input space are represented in Figure 2. The results are reported in Table 5.3, where we note that the final groups contain age values that are actually contiguous.

Table 5.1: I stage- Clusters of ages.

Age index	Ages	Relative exposures
1	18 - 22	0.049050
2	23 - 25	0.065892
3	26 - 27	0.049539
4	28 - 29	0.052641
5	30 - 31	0.055568
6	32 - 33	0.051953
7	34 - 35	0.050169
8	36 - 38	0.069945
9	39 - 40	0.045085
10	41 - 42	0.045359
11	43 - 45	0.064289
12	46 - 47	0.044006
13	48 - 50	0.060712
14	51 - 53	0.051452
15	54 - 56	0.052414
16	57 - 59	0.044891
17	60 - 62	0.037208
18	63 - 66	0.042976
19	67 - 71	0.037110
20	72 -	0.029742

Table 5.2: I stage – Clusters of claim frequencies.

Claim-freq. index	Ages	Centroids	Relative exposures
1	35-41 43 64 77 81 85 88 89 91 93 94	0.082046	0.197729
2	31-33 42 61 68 69 90 95	0.089643	0.130210
3	30 34 46 50 56 63 65-67 72 79 80	0.093848	0.158257
4	29 44 47 48 51 59 60 62 70 73 75 76 78	0.100367	0.166243
5	27 28 45 53-55 57 58 71 74 84	0.106366	0.167085
6	49 52 83	0.114182	0.040280
7	23-26 87	0.135333	0.090205
8	21 82	0.157709	0.016459
9	18-20 22 86 92	0.188920	0.033533

To appreciate the features of the resulting groups we compare in Figure 3 the original claim frequencies with the centroids of the clusters.

Incidentally, observe that young drivers show a quite high risk level and in fact the claim frequency progressively decreases in clusters 2, 3 and 4, whereas clusters 5 and 6 show a higher risk level again. This is a well-known phenomenon present in the Italian market and it is explained by the fact that in the age classes of insured 44-59 we find the insured whose young sons or daughters get their driving licence and begin to drive their parents'car.

Table 5.3: II stage – Clusters of ages of the insured.

Cluster	Ages	Claim-frequency centroids	Relative exposure
1	18 - 26	0.1507	0.139073
2	27 - 29	0.1056	0.078049
3	30 - 35	0.0904	0.157690
4	36 - 43	0.0835	0.181334
5	44 - 50	0.1011	0.148062
6	51 - 59	0.1042	0.148756
7	60 -	0.0945	0.147036

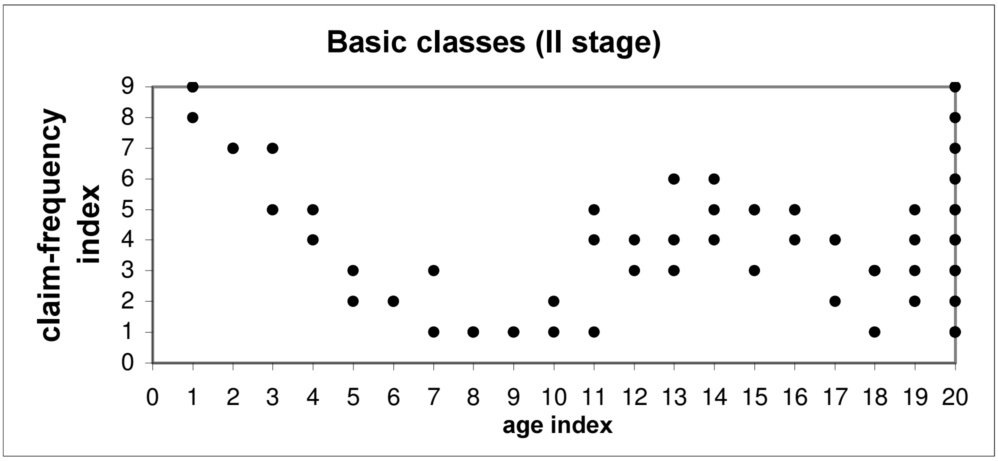

Figure 2

As far as the goodness of the clustering is concerned, in this case the distortion cannot be considered an acceptable criterion to choose among different groupings. In fact, in this application the distortion calculated from the centroids of the clusters is 1.2026 x 10^{-4}, an extremely high level when compared with distortion values of the subdivisions in 7 clusters obtained in Section 4. Clearly, the continuity of the elements in the groups is a valuable result but it cannot be evaluated by means of this traditional measure.

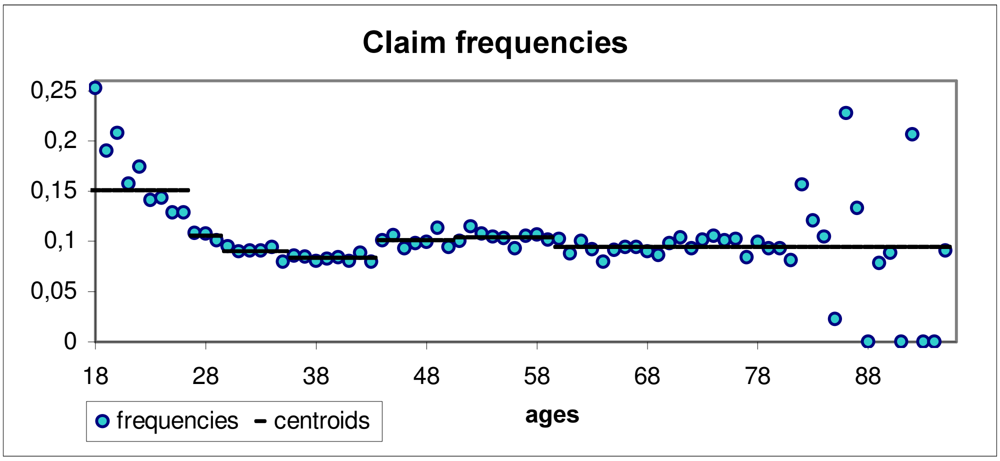

Figure 3

References

Anderberg, M. R., (1973): *Cluster Analysis for applications.* Academic Press, New York, N.Y.

Bishop, C.M., (1995): *Neural networks for pattern recognition.* Clarendon Press, Oxford

Black, J.V., (1992): A Unified Framework for Vector Quantisers, *Defence Research Agency Malvern Memorandum 4670*

de Bodt, E., Cottrell, M., and Verleysen, M., (1999): Using the Kohonen Algorithm for Quick Initialization of Simple Competitive Learning Algorithm, *Proceedings of ESANN 99*, D-Facto publications

de Bodt, E., Verleysen, M., and Cottrell, M., (1997): Kohonen maps versus vector quantization for data analysis, *Proceedings of ESANN 97*, D-Facto publications

Cottrell, M., Fort, J. C., and Pagès G., (1998): Theoretical Aspects of the SOM Algorithm, *Neurocomputing*, 21 No.1-3,119-138

Dickmann, H., (1978): Einsatz der Clusteranalyse bei Klassifikationsproblemen in der Versicherungswirtschaft, *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, XIII (4), 387-401

van Eeghen, J., Greup, E. K., and Nijssen, J. A., (1983): Rate making, *Surveys of Actuarial Studies*, 2, Nationale-Nederlanden N.V.

Giulini, S., Pelessoni, R., and Picech, L., (1997): Determination of tariff classes: cluster analysis methods and unsupervised neural networks, *Proceedings of the XXVIII International Astin Colloquium*, 129-150, Cairns (Australia)

Hassoun, M.H., (1995): *Fundamentals of artificial neural networks*, The MIT Press, London

Hertz, J., Krogh, A., and Palmer, R.G., (1991): *Introduction to the theory of neural computation*, Addison-Wesley, Redwood City, California

Jain, A. K., and Farrokhnia, F., (1991): Unsupervised Texture Segmentation Using Gabor Filters, *Patter Recognition*, 24 No.12, 1167-1186

Kohonen, T., (1984): *Self-Organization and Associative Memory*, Springer, Berlin

Kohonen, T., (1995): *Self-Organizing Maps*, Springer, Berlin.

Linde, Y., Buzo, A., and Gray, R., (1980): An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications*, COM-28, 84-99

Loimaranta, K., Jacobsson, J., and Lonka, H., (1980): On the use of mixture models in clustering multivariate frequency data, *Transactions of the 21st International Congress of Actuaries*, 2, 147-161

Lowe, J.A., and Pryor, L.M., (1996): Neural Networks v. GLMs in pricing general insurance, *General Insurance Convention*, 2, 417-438

Luttrell, S.P., (1990): Derivation of a Class of Training Algorithms, *IEEE Transactions on Neural Networks*, 1 No.2, 229-232

Pelessoni R., and Picech, L., (1997): Simple Competitive Learning, Kohonen S.O.M. and Schmitter-Straub's Method: a Proposal in Finding "Good" Tariff Classes, *Atti della Giornata di Studio su "Aspetti Scientifici e Didattici nella Teoria del Rischio"* , Campobasso, 121-142

Ripley, B.D., (1996): *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge

Schmitter H., and Straub, E., (1975): How to find the right subdivision into tariff classes, *Astin Bulletin*, 8 No.2, 257-263