

# A Neural Network Approach for Credit Risk Evaluation

Eliana Angelini

*Author's Affiliation*

Giacomo di Tollo

*Author's Affiliation*

Andrea Roli

*Author's Affiliation*

**Abstract.** The Basel Committee on Banking Supervision proposes a capital adequacy framework that allows banks to calculate capital requirement for their banking books using internal assessments of key risk drivers. Hence the need for systems to assess credit risk. Among the new methods, artificial neural networks have shown promising results. In this work, we describe the case of a successful application of neural networks to credit risk assessment. We developed two neural network systems, one with a standard feedforward network, while the other with a special purpose architecture. The application is tested on real-world data, related to Italian small businesses. We show that neural networks can be very successful in learning and estimating the *bonis/default* tendency of a borrower, provided that careful data analysis, data pre-processing and training are performed.

**Keywords:** credit risk, Basel II, neural networks

## 1. Introduction

The Basel Committee on Banking Supervision, with its revised capital adequacy framework “International Convergence of Capital Measurement and Capital Standards” (Basel Committee on Banking Supervision, 2005) - commonly known as Basel II - proposes a more flexible capital adequacy framework to encourage banks to make ongoing improvements in their risk assessment capabilities . The text reflects the results of extensive consultations with supervisors and bankers worldwide<sup>1</sup>. Regulators allow banks the discretion to calculate capital requirement for their banking books using “internal assessments” of key risk drivers, rather than the alternative regulatory standardized model: the risk weights and thus capital charge are determined through the combination of quantitative inputs provided by bank and formulas

---

<sup>1</sup> The Committee subsequently released the first round of proposals for revising the capital adequacy framework in June 1999; than in January 2001 and April 2003 and furthermore conducted three quantitative impact studies related to its proposals. As a results of these labours, many valuable improvements have been made to the original proposals.



specified by the Committee. For the first time, banks will be permitted to rely on their own assessments of a borrower's credit risk.

Credit risk has long been an important and widely studied topic in bank lending decisions and profitability. For all banks, credit remains the single largest risk, difficult to offset, despite advances in credit measurement techniques and the diversification of portfolio. Continuing increases in the scale and complexity of financial institutions and in pace of their transactions demand that they employ sophisticated risk management techniques and monitor rapidly changing credit risk exposures. At the same time, fortunately, advances in information technology have lowered the cost of acquiring, managing and analysing data, in an effort to build more robust and sound financial systems.

In recent years, a number of the largest banks have developed sophisticated systems in an attempt to assess credit risk arising from important aspects of their business lines. What have been the benefits of the new model-based approach to risk measurement and management? The most important is that better risk measurement and management contribute to a more efficient capital allocation<sup>2</sup>. When risk is better evaluated, it can be more accurately priced and it can be more easily spread among a larger number of market participants. The improvement in credit risk modelling has led to the development of new markets for credit risk transfer, such as credit derivatives and collateralised debt obligations (CDOs). These new markets have expanded the ways that market participants can share credit risk and have led to more efficient pricing of that risk (R.W.Ferguson, 2001).

The aim of this paper is to apply the neural network approach in the small-business lending analysis to assess credit risk of listed companies. The focus of this article is on the empirical approach, by using two different architectures of neural networks trained on real-world data.

The paper is organized as follows. The paper begins (section 2) by stating the overall objectives of the Basel Committee in addressing the topic of sound practices to the credit risk management process. Moreover, the paper presents important elements of regulatory function. Section 3 presents an analysis of the conceptual methodologies to credit risk modelling and focuses on the various techniques used for parameter estimation. We have tried to simplify the technical details and analytics surrounding these models. Finally, we emphasize on neural

---

<sup>2</sup> "...More accurate risk measurement and better management do not mean the absence of loss. Those outcomes in the tails of distributions that with some small probability can occur, on occasion do occur; but improved risk management has meant that lender and investors can more thoroughly understand their risk exposure and intentionally match it to their risk appetite, and they can more easily hedge unwanted risk..." (R.W.Ferguson, 2001)

network models, extremely useful for analysing small-business lending problems. In section 4 we first briefly overview the main principles and characteristics of neural networks; then we describe the models we developed and tested. In section 5 the data set used is illustrated: We select 30 companies as samples and assign each of them to one of the two groups: a “good” one, which means that the economic and financial situation is good, and a “bad” one, which means that the company is close to default. Section 6 discusses the experimental settings and the results we obtained, leading to considerable accuracy in prediction. The paper concludes with a discussion of advantages and limitations of the solution achieved and the future work for improvement.

## 2. Key Elements of the New Basel Capital Accord

### 2.1. WHY NEW CAPITAL REQUIREMENTS?

In the last analysis, the Basel Supervisor’s Committee has published its document describing the proposal in its near-final form. The industry demonstrated an actual support for this survey and 265 banks from nearly 50 countries concrete assessment of how the Committee’s proposals will functioning as we had hoped. Committee efforts are geared for improving risk management, measuring risk more accurately; communicating those measurements to the management, to supervisors, and to the public; and, of course, relating risk both to capital requirement and to supervisory focus (R.W.Ferguson, 2003).

The existing rules, based on relatively simple 1988 Basel Accord, represented an important step in answering the age-old question of how much capital is enough for banks to hedge economic downturns. Specifically, under the current regulatory structure, virtually all private-sector loans are subject to the same 8 percent capital ratio with no account of the size of the loan, its maturity and, most importantly, the credit quality of the borrower. Thus, loans to a firm near bankruptcy are treated (in capital requirement terms) in the same fashion as loans to a AAA borrowers. Moreover, the current capital requirement is additive across all loans; there is no allowance for lower capital requirements because of a greater degree of diversification in the loan portfolio (A.Saunders, 1999).

By the late 1990s, it became clear that the original Accord was becoming outdated. Its nature has had a tendency to discourage certain types of bank lending. It has also tended to encourage transactions whose sole benefits is regulatory capital relief (W.J.McDonough, 2003). Further, the international competition, the globalisation and the improvements in the risk management tools, changed the way that banks

monitor and measure risk in manner that the 1988 Accord could not anticipate. To response to these challenges, the Committee began a few years ago to develop a more flexible capital adequacy framework. The New Accord consists of three pillars: minimum capital requirements, supervisory review of capital adequacy and public disclosure. The Committee believes that all banks should be subject to a capital adequacy framework comprising minimum capital requirements, supervisory review, and market discipline.

The objective is reached by giving banks a range of increasingly sophisticated options for calculating capital charges. Banks will be expected to employ the capital adequacy method most appropriate to the complexity of their transactions and risk profiles. For credit risk, the range of options begins with the standardized approach and extends to the internal rating-based (IRB) approach. The standardized approach is similar to the current Accord: banks will be expected to allocate capital to their assets based on the risk weights assigned to various exposures. It improved on the original Accord by weighting those exposures based on each borrower's external credit risk rating. Clearly, the IRB approach is a major innovation of the New Accord: bank internal assessments of key risk drivers are primary inputs to the capital requirements. For the first time, banks will be permitted to rely on their own assessments of a borrower's credit risk. The close relationship between the inputs to the regulatory capital calculations and banks' internal risk assessments will facilitate a more risk sensitive approach to minimum capital. Changes in a client's credit quality will be directly reflected in the amount of capital held by banks.

How will the New Basel Accord promote better corporate governance and improve risk management techniques?

First, the Basel Committee has expressly designed the New Accord to provide tangible economic incentives for banks to adopt increasingly sophisticated risk management practices. Banks with better measure of their economic risks will be able to allocate capital more efficiently and more closely in line with their actual sensitivity to the underlying risks. Second, to achieve those capital benefits, the more advanced approaches to credit and operational risk require banks to meet strong process control requirements. Again, the increasing focus on a bank's control environment gives greater weight to the management disciplines of measuring, monitoring and controlling risk (W.J.McDonough, 2003).

## 2.2. "CORE COMPONENTS" IN THE IRB APPROACH

In the IRB Approach to credit risk there are two variants: a foundation version and an advanced version. In the first version, banks

must provide internal estimates of probability of default (PD)- which measures the likelihood that the borrower will default over a given time horizon. In addition, in the advanced approach, banks, subject to certain minimum conditions and disclosure requirements, can determine other elements needed to calculate their own capital requirements. They are: (a) Loss given default (LGD), which measures the proportion of the exposure that will be lost if a default occurs; (b) Exposure at default (EAD), which for loan commitments measures the amount of the facility that is likely to be drawn if a default occurs; (c) Maturity (M), which measures the remaining economic maturity of the exposure (Basel Committee on Banking Supervision, 2005). In other words, the two approaches differ primarily in terms of the inputs that are provided by the bank based on its own estimates and those that have been specified by the Committee. The risk weights and thus capital charges are determined through the combination of quantitative inputs provided by banks and formulas specified by the supervisor. The risk weight functions have been developed for separate asset classes. For corporate, sovereign and banks formulas are:

$$\text{Correlation}(R) = 0.12 \times \frac{1 - e^{-50 \times \text{PD}}}{1 - e^{-50}} + 0.24 \times \left[ 1 - \frac{1 - e^{-50 \times \text{PD}}}{1 - e^{-50}} \right];$$

$$\text{Maturity adjustment}(b) = (0.11852 - 0.05478 \text{ times } \ln(\text{PD}))^2;$$

$$\begin{aligned} \text{Capital requirement}(K) = & \\ & \left[ \text{LGD} \times N \left[ (1 - R)^{-0.5} \times G(\text{PD}) + \left( \frac{R}{1 - R} \right)^{0.5} \times G(0.999) \right] - \text{PD} \times \text{LGD} \right] \times \\ & \times (1 - 1.5 \times b)^{-1} \times (1 + (M - 2.5) \times b); \end{aligned}$$

$$\text{Risk-weighted assets (RWA)} = K \times 12.5 \times \text{EAD}.$$

Where:  $N(x)$  denotes the cumulative distribution function for a standard normal random variable (i.e. the probability that a normal random variable with mean zero and variance of one is less than or equal to  $x$ ).  $G(z)$  denotes the inverse cumulative distribution function for a standard normal random variable (i.e. the value of  $x$  such that  $N(x) = z$ ). PD and LGD are measured as decimals, and EAD is measured as currency, except where explicitly noted otherwise. This risk weight function, based on modern risk management techniques, translates a bank's inputs into a specific capital requirement. Under the IRB approach for corporate credits, banks will be permitted to separately distinguish exposures to small and medium sized entities (SME). They are defined as corporate exposures where the reported sales for the

consolidates group of which the firm is part is less than €50 million from those to large firms. A firm size adjustment  $[0.04 \times 1 - (S - 5)/45]$  is made to the corporate risk weight formula for exposures to SME borrowers (Basel Committee on Banking Supervision, 2005).

$$\text{Correlation}(R) = 0.12 \times \frac{1 - e^{-50 \times \text{PD}}}{1 - e^{-50}} + 0.24 \times \left[ 1 - \frac{1 - e^{-50 \times \text{PD}}}{1 - e^{-50}} \right] - 0.04 \times (1 - (S - 5)/45).$$

$S$  is expressed as total annual sales in millions of euros with value of  $S$  falling in the range of equal to or less than €50 million or greater than or equal to €5 million. Another major element of the IRB approach is the treatment of credit risk mitigants, namely, collateral, guarantees and credit derivatives; in particular the LGD parameter provides a great deal of flexibility to assess the potential value of credit risk mitigation techniques. In these formulas we can see that the most important risk component is the probability of default. PD estimates must be a long-run average of one year realised default rates for borrowers in the grade. For corporate and bank exposures, the PD is the greater of the one year PD associated with the internal borrower grade to which that exposure is assigned, or 0.03%. Banks may use one or more of the three specific methods (internal default experience, mapping to external data and statistical default models) as well as other information and techniques as appropriate to estimate the average PD for each rating grade. Improvements in the rigour and consistency of credit risk measurement, the flexibility of models in responding to changes in the economic environment and innovations in financial products may produce estimates of credit risk that better reflect the credit risk of exposure. However, before a modelling approach could be used in the formal process of setting regulatory capital requirements for credit risk, regulators would have to be confident not only that models are being used to actively manage risk, but also that they are conceptually sound and empirically validated. Additionally, problems concerning data limitations and model validation must be cleared before models may be implemented in the process of setting regulatory capital requirement. At present, there is no commonly accepted framework for periodically verifying the accuracy of credit risk models; it is important to note that the internal environment in which a model operates – including the amount of management oversight, the quality of internal controls, the rigour of stress testing, the reporting process and other traditional features of credit culture – will also continue to play a key part in the evaluation of bank's risk management framework (Basel Committee on Banking Supervision, 1999).

### 3. Conceptual approaches to credit risk modelling

Over the last decade, enormous strides have been made in the art and science of credit risk measurement. Banks have devoted increased attention to measuring credit risk and have made important gains, both by employing innovative and sophisticated risk modelling techniques and also by strengthening their more traditional practices (L.H.Meyer, 2000). Measuring credit risk accurately allows banks to engineer future lending transactions, so as to achieve targeted return/risk characteristics. However, credit risk models are not a simple extension of their market risk counterparts for two key reasons (Basel Committee on Banking Supervision, 1999).

- The specification of the process of default and rating migration is severely constrained by a lack of data on the historical performance of loans and other modelled variables; most credit operations are not market to market and the predictive nature of a credit risk model does not derive from a statistical projection of future prices based on a comprehensive record of historical prices. The difficulties in specification are exacerbated by the longer term time horizons used in measuring credit risk, which suggest that many years of data, spanning multiple credit cycles, may be needed to estimate the process of default. Even if individual default probabilities could be modelled accurately, the process of combining these for a portfolio might still be hampered by the scarcity of data with which to estimate reliably the correlations between numerous variables. Hence, in specifying models parameters, credit risk models require the use of simplifying assumptions and proxy data.
- The validation of credit risk models is fundamentally more difficult than the back testing of market risk models. Where market risk models typically employ a horizon of a few days, credit risk models generally rely on a time frame of one year or more; the longer holding period, coupled with the higher confidence intervals used in credit risk models, presents problems to model-builders in assessing the accuracy of their models; the effect of modelling assumptions on estimates of the extreme tails of the distributions is not well understood.

The new models – some publicly available and some partially proprietary – try to offer “internal model” approaches to measure the credit risk of a loan or a portfolio of loans. In this section, we do not propose to make a taxonomy of these approaches, but aim to discuss key elements of the different methodologies.

First, within the current generation of credit risk models, banks employ either of two conceptual definitions of credit loss, the *default mode* (DM) paradigm or the *mark to market* (MTM) paradigm. In the first paradigm a credit loss arises only if a borrower defaults within the planning horizon<sup>3</sup>. In the absence of a default event, no credit loss would be incurred. In the case that a client defaults, the credit loss would reflect the difference between the bank's credit exposure and the present value of future net recoveries. In contrast to the DM paradigm, in the MTM models a credit loss can arise in response to the deterioration in an asset's credit quality. Given the rating transition matrix associated with each client, Monte Carlo methods are generally used to simulate migration paths for each credit position in the portfolio<sup>4</sup>. Second, there are different methodologies for the unconditional and conditional models. Unconditional approaches typically reflect customer or facility-specific information<sup>5</sup>. Such models are currently not designed to capture business cycle effects, such as the tendency for internal ratings to improve (deteriorate) more during cyclical upturns (downturns). Instead, conditional models incorporate information on the state of the economy, such as levels and trends in indicators of economic and financial health, in domestic and international employment, in stock prices and interest rates, ecc. In these models, rating transition matrices are increased likelihood of an upgrade during an upswing in a credit cycle and vice versa<sup>6</sup>.

Finally, there are different techniques for measuring the interdependence of factors that contribute to credit losses. In measuring credit risk, the calculation of a measure of the dispersion of credit risk requires consideration of the dependencies between the factors determining credit-related losses, such as correlations among defaults or rating migrations, LGDs and exposures, both for the same borrower and among different borrowers (Basel Committee on Banking Supervision, 2000).

---

<sup>3</sup> The choice of a modelling horizon of one year reflects the typical interval over which: (a) new capital could be raised; (b) new customer information could be revealed; (c) loss mitigation actions could be undertaken; (d) internal budgeting, capital planning and accounting statements are prepared; (e) credits are normally reviewed for renewal.

<sup>4</sup> See J.P.Morgan's Credit Metrics™ framework (1997) for MTM approach; Credit Risk Plus™ of Credit Suisse Financial Product (1997) for DM approach.

<sup>5</sup> Examples are CreditMetrics™ and Credit Risk +™; these modelling frameworks derive correlation effects on relationship between historical defaults and borrower-specific information such as internal risk ratings. The data is estimated over (ideally) many credit cycles.

<sup>6</sup> One example is McKinsey and Company's Credit Portfolio View™.



In closing, the fundamental elements of credit risk measurement are easy to describe in the abstract but are far more difficult to apply case by case. Each situation is unique, built around the roles and capabilities of individuals, the technology system, activities and objectives of the institutions. Anyway, to remain competitive, institutions must adapt and constantly improve their credit risk measurement techniques.

#### 4. The neural network approach

Neural networks have recently emerged as an effective method for credit scoring (Wu and Wang, 2000; Rong-Zhou et al., 2002; Pang et al., 2002; Piramuthu, 1999; Atiya, 2001). They differ from classical credit scoring systems, such as the *Z-score model* (Altman, 1968), mainly in their *blackbox* nature and because of they assume a non-linear relation among variables. Neural networks are learning systems which can model the relation between a set of inputs and a set of outputs, under the assumption that the relation is nonlinear. They are considered blackboxes, since, in general, it is not possible to extract symbolic information from their internal configuration.

In this section, we introduce neural networks with the aim of defining the background needed to discuss the model we propose and the experimental results.

##### 4.1. NEURAL NETWORKS: AN INTRODUCTION

Neural networks are machine learning systems based on a simplified model of the biological neuron (S.Hykin, 1999). In the same way as the biological neural network changes itself in order to perform some cognitive task (such as recognizing faces or learning a concept), artificial neural networks modify their internal parameters in order to perform a given computational task.

Typical tasks neural networks perform efficiently and effectively are: Classification (i.e., deciding which category a given example belongs to), recognizing patterns in data, prediction (such as the identification of a disease from some symptoms, or the identification of causes, once effects are given).

The two main issues to be defined in a neural network application are the network typology and structure and the learning algorithm (i.e., the procedure used to adapt the network so as to make it able to solve the computational task at hand).

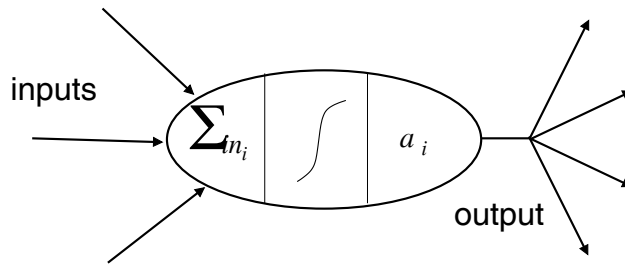


Figure 1. Artificial neuron basic structure.

An artificial neural network<sup>7</sup> is composed of a bunch of neurons, connected in a predefined topology. There are some possible topologies, usually depending on the task the network has to learn. Usually, the network topology is kept constant, but in some applications (for instance in robotics) the topology itself can be considered as a parameter and can dynamically change. The connections (links) among neurons have associated a *weight* which determines type and intensity of the information exchanged. The set of weights represents, in essence, the information the network uses to perform the task, i.e., given a topology, the weights represent the *functions* that defines the network behavior.

The artificial neuron is an extremely simplified model of the biological neuron and it is depicted in Fig.1. Neurons are the elementary computational units of the network. A neuron receives inputs from other neurons and produces an output which is transmitted to other destination neurons.

The generation of the output is divided in two steps. In the first step, the weighted sum of inputs is evaluated, i.e., every single input is multiplied by the weight on the corresponding link and all these values are summed up. Then, the *activation* is evaluated by applying a particular *activation function* to the weighted sum of inputs. In formulas, for neuron  $i$  receiving inputs from neurons in the set  $I$ ,

$$y_i = \sum_{j \in I} W_{j,i} a_j \quad \text{input evaluation}$$

$$a_i = g(y) \quad \text{the activation function } g \text{ is applied}$$

where  $W_{j,i}$  is the weight of the link connecting neuron  $j$  with neuron  $i$ , and  $a_j$  is the activation of neuron  $j$ .

Some kinds of activation function are used, for instance: linear, step function and sigmoid (see Fig.2).

The most used network topologies are the following:

---

<sup>7</sup> In the following, we will skip the adjective “artificial”, since we will not deal with biological neural networks.

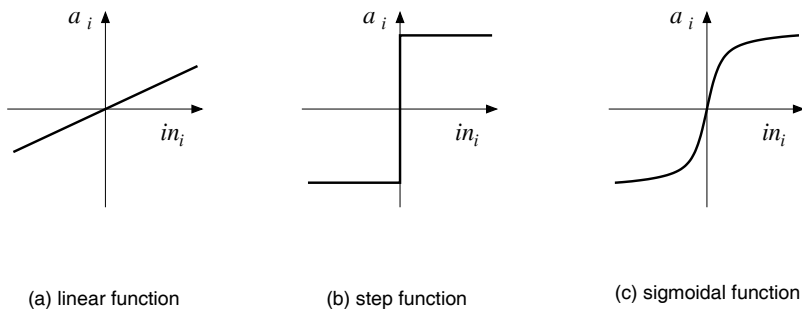


Figure 2. Examples of activation functions.

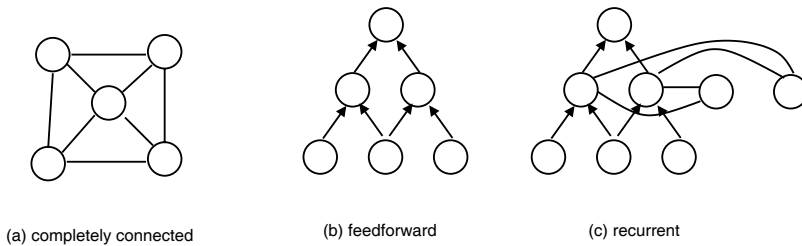


Figure 3. Examples of network topologies.

- Layered
- Completely connected

Networks of the first category have neurons subdivided in layers. If the connections are only in one direction (i.e., each neuron receives inputs from the previous layer and sends output to the following layer), they are called *feedforward networks*. Otherwise, if also 'loops' are allowed, the network is called *recurrent network*. Completely connected networks, on the other hand, have neurons which are all connected with each other<sup>8</sup>. In Fig.3, examples of the three main kinds of topologies are depicted.

Before the neural network can be applied to the problem at hand, a specific tuning of its weights has to be done. This task is accomplished by the learning algorithm which trains the network and iteratively modifies the weights until a specific condition is verified. In most applications, the learning algorithm stops as soon as the discrepancy (*error*) between desired output and the output produced by the network falls

<sup>8</sup> In principle, a completely connected network can be seen as a special case of a recurrent network with one level of  $n$  neurons and  $n^2$  connections.

below a predefined threshold. There are three typologies of learning mechanisms for neural networks:

- Supervised learning
- Unsupervised learning
- Reinforced learning

*Supervised learning* is characterized by a *training set* which is a set of correct examples used to train the network. The training set is composed of pairs of inputs and corresponding desired outputs. The error produced by the network then is used to change the weights. This kind of learning is applied in cases in which the network has to learn to generalize the given examples. A typical application is classification: A given input has to be inserted in one of the defined categories.

In *unsupervised learning* algorithms, the network is only provided with a set of inputs and no desired output is given. The algorithm guides the network to self-organize and adapt its weights. This kind of learning is used for tasks such as data mining and clustering, where some regularities in a large amount of data have to be found.

Finally, *reinforced learning* trains the network by introducing prizes and penalties as a function of the network response. Prizes and penalties are then used to modify the weights. Reinforced learning algorithms are applied, for instance, to train adaptive systems which perform a task composed of a sequence of actions. The final outcome is the result of this sequence, therefore the contribution of each action has to be evaluated in the context of the action chain produced.<sup>9</sup>

Diverse algorithms to train neural networks have been presented in the literature. There are algorithms specifically designed for a particular kind of neural networks, such as the *backpropagation algorithm* (Werbos, 1988) or general purpose algorithms, such as *genetic algorithms* (Mitchell, 1998) and *simulated annealing* (Kirkpatrick et al., 1983).

At the end of this introductory section, we would like to remark advantages and limits of systems based on neural networks. The main advantages have to be found in their learning capabilities and in the fact the derived model does not make any assumption on the relations among input variables. Conversely, a theoretical limit of neural networks is that they are *black-box* systems and the extraction of symbolic knowledge is awkward. Moreover, design and optimization neural network

---

<sup>9</sup> In these cases it is difficult, if not impossible, to produce an effective training set, hence the need of the use of prizes and penalties, instead of the definition of an explicit error as in the case of supervised learning.

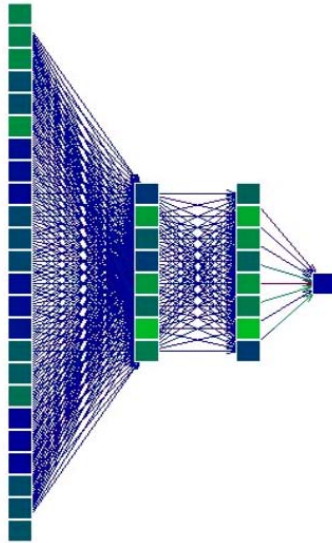


Figure 4. A four layers feedforward network.

methodologies are almost all empirical, thus the experience and sensibility of the designer have a strong contribution in the final success. Nevertheless, with this work we show that some useful general design and parameters optimization guidelines exist.

In the next section we will describe in more detail the neural network models we adopted. We used a classical feedforward network and a variation of feedforward network with *ad hoc* connections. Both networks are trained with the backpropagation algorithm.

#### 4.2. OUR MODELS

In our experiments, we used a feedforward neural network –in the classical topology– and with a feedforward neural network with *ad hoc* connections.

The feedforward network architecture is composed of an input layer, two *hidden* layers and an output layer (composed of only one neuron). In Fig.4 an example of a classical two hidden layers feedforward network is represented. In the following, we will indicate with  $W_{k,j}$  the weights between input and hidden neurons and with  $W_{j,i}$  the weights between hidden and output neurons. Moreover, the activation will be  $I_k$ ,  $H_j$  and  $O_i$  for input, hidden and output neurons respectively.

The feedforward network with *ad hoc* connections (thereinafter referred to as *ad hoc* network) is a four layers feedforward network with the input neurons grouped by three. Each group is connected to one

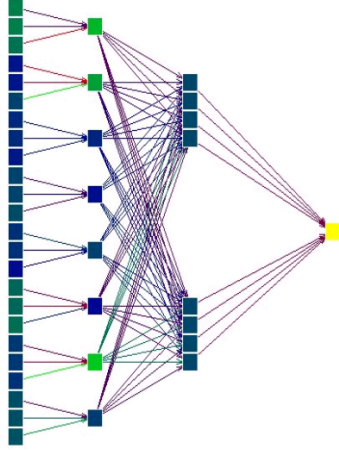


Figure 5. The ad hoc network used in our experiments.

neuron of the following layer. The reasons for choosing this topology are to be found in the actual data we used and will be described in the next section. The ad hoc network topology is depicted in Fig.5.

Both the networks have been trained by means of a supervised algorithm, namely the backpropagation algorithm. This algorithm performs an optimisation of the network weights trying to minimize the error between desired and actual output. For each output neuron  $i$ , the error is:  $Err_i = T_i - O_i$ , where  $T_i$  is the desired output. The weight update formula used to change the weights  $W_{j,i}$  is the following:

$$W_{j,i} \leftarrow W_{j,i} + \eta H_j Err_i g'(in_i),$$

where  $\eta$  is a coefficient which controls the amount of change (the *learning rate*),  $g'$  is the first derivative of the activation function  $g$  and  $in_i$  is equal to  $\sum_j W_{j,i} H_j$ .

The formula for updating weights  $W_{k,j}$  is similar:

$$W_{k,j} \leftarrow W_{k,j} + \eta I_k \Delta_j,$$

where  $I_k$  is the  $k$ -th input<sup>10</sup> and  $\Delta_j = g'(in_j) \sum_i W_{j,i} Err_i g'(in_i)$ .

The algorithm is iterated until a stopping criterion is satisfied<sup>11</sup>.

The definition of both the network structure and the algorithm needs the careful choice of parameters, such as the number of neurons at each layer and the value of the learning rate. Moreover, both the data

<sup>10</sup> The activation of input neurons is simply the input value.

<sup>11</sup> E.g., the overall network error is below a predefined threshold or the maximum amount of allotted time is reached.

used to train the network (the training set) and the data on which the performance of the network is validated (the test set) need to be carefully chosen. In Sec.6, we will describe our choices along with the benchmark used. In the next section we will describe the data set used in the experiments.

## 5. Data set

For our experiments, we used data of 76 small businesses from a bank in Italy. For each business we have data across three years (2001-2003). For each business we have 11 fields: 8 of them are financial ratios drawn from the balance sheet of the firms: The ratios (fields) on the sample set are the following:

1.  $\frac{\text{Cash flow}}{\text{Total debt}}$ ,
2.  $\frac{\text{Sales}}{\text{Stock value}}$
3.  $\frac{\text{Short-term liability}}{\text{Sales}}$
4.  $\frac{\text{Equity}}{\text{Total assets}}$ ,
5.  $\frac{\text{Financial costs}}{\text{Total debts}}$ ,
6.  $\frac{\text{Circulating capital}}{\text{Total assets}}$ ,
7.  $\frac{\text{Consumer credits}}{\text{Sales}}$ ,
8.  $\frac{\text{Value added}}{\text{Total assets}}$ ,

The remaining 7 are calculated analysing the credit positions with the supplying (“*Andamentale*” ratios) bank

9.  $\frac{\text{Utilized credit line}}{\text{Accorded credit line}}$ ,
10.  $\frac{\text{Outstanding debt (quantity)}}{\text{S.B.F. effects}}$

$$11. \frac{\text{Outstanding debt (value)}}{\text{S.B.F. effects (value)}}.$$

and with the overall Italian Banking System ( “*Centrale dei Rischi*”<sup>12</sup> ratios).

$$12. \frac{\text{Transpassing short-term}}{\text{Accorded credit line short-term}};$$

$$13. \frac{\text{Transpassing medium-long term}}{\text{Accorded credit line M-L term}};$$

$$14. \frac{\text{Utilized credit line short-term}}{\text{Accorded credit line short-term}};$$

$$15. \frac{\text{Utilized credit line M-L term}}{\text{Accorded credit line M-L term}}.$$

The sample businesses are categorized in two groups: the “bonis” group (composed of firms repaying the loan obligation at the end of the analysing period) and the “default” group (composed of firms not repaying the loan obligation at the end of the analysing period). Before feeding the neural net with our data we decided to operate some pre-processing operations. Here below we describe the issues about data-pre-processing.

5.0.0.1. *Missing and wrong values* Some values were missing from the data about firms. This occurrence can be due to two different reasons:

1. the value is missing because in that year there is no value in the data-base for that firm in that particular field, or because the value is not in the theoretical allowed range<sup>13</sup>;
2. the value is missing because of a computation-error. This happens because the variables we are working on are mathematical ratios, and in our sample some ratios can derive by a division by zero.

The usual way of overcoming this problem is to discard from the data set the all the entries of the corresponding firm (this is the approach pursued on several works about neural nets and credit risk), but operating in that way we would loss a significant amount of information. In order to preserve this information we will perform the replacement of missing values with other values. We decide to handle differently these two situations.

<sup>12</sup> This data are drawn from the “Centrale Dei Rischi”, a “Banca D’Italia” kept data-base.

<sup>13</sup> Such a kind of errors may occur because of typos and transcription errors accidentally introduced by bank employees. This situation is quite common and constitutes one of the sources of noise in real-world data.



1. In the first case (missing because of none or wrong value) we decide to substitute the empty space with the arithmetical mean of the field, being the mean calculated as the mean of the existing values belonging to that field for all the businesses in the overall period of collecting (column substitution);
2. In the latter case we decide to replace the missing value with the upper limit of the Normalization interval (see below). This choice can easily understood: as this occurrence is given by a division by zero, we can imagine the result of this ratio being  $\infty$ . If we normalize the belonging field, this occurrence will be replaced by the maximum value in the range. We mentioned earlier in our experiment this upper limit being 1.

5.0.0.2. *Erasing useless fields* We discussed about replacing missing and wrong values with suitable ones in order to allow the net using those data and preserving useful information. This operation can be harmful if the processed field show too many missing and wrong values: replacing them would convey the net to draw wrong conclusions about the variables-dynamic. Furthermore, as we will see, the overall-sample splitting into *training* and *test* set is made using some percentage ratios. For this purpose we will use the ratios [70% 30%]. In the worst case, if for a specific field there are more than 30% of missing and wrong values (on the overall observations) they can be all included in the test set. This will result in an incredible loss of generalization capability of the net. For this reason we decided not to use the fields containing more than 30% of missing and wrong values and precisely the following fields belonging to “Andamentale” and “Centrale dei Rischi” ratios.

5.0.0.3. *Data Normalization* Data normalization must be performed in order to feed the net with data ranging in the same interval for each input node. We choose to use the interval [0, 1] for each input node. The most common way for normalizing data is the *Min-Max linear transformation to [0, 1]*<sup>14</sup>, but we cannot use this formula because there are “outliers”<sup>15</sup> in the data-base we are using. Using the Min-Max formula we would lose a lot of useful information and several fields would have almost all data close to one of the limit of normalization. For this reason we decided to use the logarithmic formula to normalize data. This formula is the most flexible because can be defined by the

---

<sup>14</sup> The general formula of this kind of normalization is  $\bar{x}_{ia} = \frac{x_{ia} - \min_a}{\max_a - \min_a}$ , where  $\bar{x}_{ia}$  is the after-normalizing value and  $x_{ia}$  is the actual value.

<sup>15</sup> Outlier means that the value is too different from the other values in the field we are analysing.

user, provided that the argument of the formula being  $< 1$ . For our ratios we used the following formula:

$$\bar{x} = \log_m(x + 1)$$

Where  $m$  is near to the actual maximum of the field we are analysing. We add 1 to the argument to avoid the value being  $> 1$ .

5.0.0.4. *Correlation analysis* In order to decide which ratios to use as input variables we performed a correlation analysis: with this operation we want to find out the most strongly correlated variables and remove them from our further experiments. The results of this analysis show that there is no strong correlation between couples of variables, so we use the 11 fields examined so far.

5.0.0.5. *Training and test set* We select 53 firms to consist of training set and 23 firms to consist of test set. We want to reserve, defining both sets, the ratio between bonis and default firms existing in the overall sample set. For this reason we use 33 bonis examples and 20 default examples to define the training set. For the same purpose we use 15 bonis examples and 8 default examples to define the test set.

## 6. Experimental Results

The two network architectures introduced in Sec.4.2 have been trained with backpropagation using training and test sets as described in the previous section.

The inputs of the networks are the eleven (normalized) attributes. In the classical feedforward network, they are simply given as an ordered array, while in the ad hoc network they are first grouped by three, each group corresponding to the values of an attribute over three years.

The output  $y$  of the network, a real value in the range  $[0, 1]$  is interpreted as follows:

- If  $y < 0.5$  then the input is classified as **bonis**;
- otherwise, the input is classified as **default**

A large number of tests has been performed trying to optimally tune parameters of the network and the algorithm. Despite many years of research in neural networks, parameter tuning is usually still lacking a theoretical framework, therefore this design phase is still performed empirically. Nevertheless, some general methodological guidelines are possible and useful. For instance, our approach is based on

a simple –but effective– systematic optimisation procedure. More formally, we applied a procedure similar to a gradient ascent algorithm in the space of parameters (Blum and Roli, 2003). First, we selected for each parameter  $p_h$  (e.g., number of hidden neurons or backpropagation parameters such as learning rate and momentum) a set of values  $V_h$  ( $h = 1, \dots, N_{param}$ ). Thus, the problem is to find an assignment  $\{(p_1, v_1), \dots, (p_{N_{param}}, v_{N_{param}})\}$ ,  $v_h \in V_h$ , that leads to a network with optimal performance. We start from a heuristically generated assignment  $\{(p_1, v_1^0), \dots, (p_{N_{param}}, v_{N_{param}}^0)\}$  and we iterate the following procedure: at step  $h$  the current assignment is  $\{(p_1, v_1^*), \dots, (p_{h-1}, v_{h-1}^*), \dots, (p_h, v_h^0), \dots, (p_{N_{param}}, v_{N_{param}}^0)\}$  (only parameters  $p_1, \dots, p_{h-1}$  have been assigned) and the network is trained on each instance for every possible value of parameter  $p_h$ , while the other parameters are kept constant. Then, the optimal value  $v_h^*$  is chosen. At the end, we obtain the assignment  $\{(p_1, v_1^*), \dots, (p_h, v_h^*), \dots, (p_{N_{param}}, v_{N_{param}}^*)\}$  that is, at least, not worse than any other partial assignment  $\{(p_1, v_1^*), \dots, (p_{h-1}, v_{h-1}^*), \dots, (p_h, v_h^0), \dots, (p_{N_{param}}, v_{N_{param}}^0)\}$ .

We performed a long series of experiments, of which we report only the ones corresponding to the networks achieving the best results in terms of classification. We obtained a very effective tuning for the classical feedforward network by applying a slight variation of the standard backpropagation algorithm that does not propagate errors if they are below a given threshold. With this technique we were able to produce a network with null error on the training set and an error of 8.6%, corresponding to only wrong bonis classifications (i.e., an input that should be classified as bonis is classified as default). We remark the fact that this network was able to correctly classifying all the default cases, that are considerably riskier than bonis ones. As in machine learning and data mining techniques is often discussed (Han and Kamber, 2000), false negatives are usually much more dramatically important in real world cases (e.g., in diagnosis). As reported in table I, the network performance is very robust with respect to the number of hidden neurons. In the table, we report the number of neurons in the hidden layer, the wrong bonis (*misbo*) and default (*misdef*) classifications and the global error (i.e., the overall error on the training/test set), in the training set and the test set respectively. All errors are reported in percentage. For completeness, we report also the values of the algorithm parameters:  $\eta = 0.2$ ,  $\beta = 0$ ,  $\delta = 0.1$ ; initial weights are randomly chosen in the range  $[-1, 1]$ .

With the ad hoc network we could also achieve very good results, even with a lower error on the test set that was equal to 4.3% in our best configuration. Nevertheless, the errors are all related to false negative

Table I. Best results achieved with a classical feedforward network. we report the number of neurons in the hidden layer,

# of hidden neurons	tr. set misbo	tr. set misdef	tr. set error	test set misbo	test set misdef	test set error
25	0%	0%	0%	13.3%	0%	8.6%
26	0%	0%	0%	13.3%	0%	8.6%
27	0%	0%	0%	13.3%	0%	8.6%
28	0%	0%	0%	13.3%	0%	8.6%
29	0%	0%	0%	13.3%	0%	8.6%
33	0%	0%	0%	13.3%	0%	8.6%

cases, i.e., the network classifies as bonis inputs that should be classified as default. Table II summarizes the performance of the best found configuration. For this reason, we may consider the two networks as complementary: The one returns *safe* answers, while having a – rather very small – error; the second has a very high performance in terms of overall error on the test set, but it wrongly classifies positive cases. The parameters used for training the ad hoc network are the following:  $\eta = 0.8$ ,  $\beta = 0.2$ ,  $\delta = 0$ ; initial weights are randomly chosen in the range  $[-1, 1]$ .

Table II. Best results achieved with the ad hoc feedforward network. We report the wrong bonis (*misbo*) and default (*mis*

# of hidden neurons	tr. set misbo	tr. set misdef	tr. set error	test set misbo	test set misdef	test set error
11 + 11 (2 layers)	0%	5%	1.8%	0%	12.5%	4.3%

## 7. Conclusion and future work

In this paper, we have presented an application of artificial neural network to credit risk assessment. We have discussed two neural architecture for the classification of borrowers into two distinct classes: bonis and default. The system has been trained and tested on data

related to Italian small businesses. One of the system we developed is based on a classical feedforward neural network, while the other one has a special purpose feedforward architecture. Results in both cases show that the approach is very effective and leads to a system able to correctly classify the inputs with a very low error. The overall performance of the networks we developed can be considered as a state-of-the-art result. One of the reasons for this performance is the careful analysis of the data at hand. In fact, real-world data are often noisy and incomplete; therefore, our analysis was aimed at eliminating wrong values and replace empty entries with meaningful values. Moreover, since also data normalization plays an important role in the final performance, we investigated some normalization procedures in order to keep as much information as possible in the inputs used to feed the network.

This empirical work proves, on the one hand, evidence for the actual applicability of neural networks in credit risk applications, especially as black-box non-linear systems to be used in conjunction with classical rating and classification systems. On the other hand, this research also shows that the critical issues for developing such a kind of systems are data analysing and processing.

Future work is focused on both methodological and application issues. As to methodology, we are currently working on the design of procedural techniques for data analysis and processing and for parameter optimisation. On the side of applications, we plan to assess the generalization capabilities of the networks we obtained by testing them on wider data bases.

## References

- Altman, E.: 1968, 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy'. *J. Finance* **13**.
- A.Saunders: 1999, *Credit Risk Measurement*. John Wiley & Sons.
- Atiya, A.: 2001, 'Bankruptcy Prediction for Credit Risk Using Neural Networks: A survey and New Results'. *IEEE Transactions on Neural Networks* **12**(4).
- Basel Committee on Banking Supervision: 1999, 'Credit risk modelling: Current Practice and applications'. mimeo.
- Basel Committee on Banking Supervision: 2000, 'Overview of Conceptual Approaches to Credit Risk Modelling, Part II'. mimeo.
- Basel Committee on Banking Supervision: 2005, 'International Convergence of Capital Measurement and Capital Standards. A revised framework'. Bank for International Settlements, Basel.
- Blum, C. and A. Roli: 2003, 'Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison'. *ACM Computing Surveys* **35**(3), 268-308.

- Han, J. and M. Kamber: 2000, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers.
- Kirkpatrick, S., C. Gelatt, and M. Vecchi: 1983, 'Optimization by simulated annealing'. *Science*, 13 May 1983 **220**(4598), 671–680.
- L.H.Meyer: 2000, 'Why risk management is important for global financial institutions?'. *BIS Review* (68).
- Mitchell, M.: 1998, *An introduction to genetic algorithms*. Cambridge, MA: MIT press.
- Pang, S., Y. Wang, and Y. Bai: 4-5 Novembre 2002, 'Credit scoring model based on neural network'. In: *Proc. of the First International Conference on Machine Learning and Cybernetics*. Beijing.
- Piramuthu, S.: 1999, 'Financial credit-risk evaluation with neural and neurofuzzy systems'. *European Journal of Operational Research* **112**.
- Rong-Zhou, L., P. Su-Lin, and X. Jian-Min: 4-5 Novembre 2002, 'Neural Network Credit-Risk Evaluation Model based on Back-Propagation Algorithm'. In: *Proceedings of the First International Conference on Machine Learning and Cybernetics*. Beijing.
- R.W.Ferguson: 2001, 'Credit risk management: models and judgment'. *BIS Review* (85).
- R.W.Ferguson: 2003, 'Basel II: A case study in Risk Management'. *BIS Review* (20).
- S.Hykin: 1999, *Neural Networks: A Comprehensive Foundation*. Prentice Hall International, Inc., second edition.
- Werbos, P.: 1988, 'BackPropagation, Past and future'. In: *Proceedings of the IEEE International Conference on Neural Networks*.
- W.J.McDonough: 2003, 'Risk Management, supervision and the New Basel Accord'. *BIS Review* (5).
- Wu, C. and X. Wang: 2000, 'A Neural Network approach for analyzing small business lending decisions'. *Review of Quantitative Finance and Accounting* **15**.