



# Uma abordagem difusa ao risco na concessão de crédito

Joana Patrícia Lameira Ferreira

Dissertação para a obtenção do Grau de

**Mestre em Métodos Quantitativos em Finanças**

## **Júri**

**Presidente:** Carlos Manuel Rebelo Tenreiro da Cruz

**Orientador:** Paulo Eduardo Aragão Aleixo e Neves de Oliveira

**Vogal:** José Maria Ruas Murteira

**Data:** Agosto de 2013



# Resumo

A concessão de crédito aos candidatos é importante para o desenvolvimento e sobrevivência das instituições financeiras que concedem crédito. No presente trabalho, iremos analisar e avaliar os níveis de risco de crédito dos candidatos através dum conjunto de critérios pré-definidos. Recorreremos ao credit scoring, para determinar fatores chave que influenciem a probabilidade de incumprimento dos candidatos, permitindo assim, a classificação dos mesmos em grupos distintos e como consequência, a decisão sobre a aceitação ou não da proposta em análise. Para fazer este estudo seguiremos um modelo GDM fuzzy. Posteriormente, iremos aplicá-lo a dados fornecidos pelo Montepio.

**Palavras Chave:** Classificação de crédito, Modelo *GDM Fuzzy*, Análise de Componentes Principais, Regressão Logística

# Abstract

Conceding credit to candidates is important to the development and survival of the financial institutions that concede credit. In this thesis, we will analyse and evaluate the levels of credit risk of our candidates using a set of criteria defined previously. We use credit scoring to determine key factors that influence the probability that the candidates fail to pay their credits, and that will allow us to sort them in different groups, in order to decide whether or not we should accept their credit proposal. On this study we will use a GDM fuzzy model. Afterwards we will apply our model to a set of data given to us by the bank Montepio.

**Keywords:** Credit Scoring, GDM Fuzzy model, Principal Component Analysis, Logistic Regression



# Agradecimentos

*A realização desta dissertação marca o fim de uma etapa importante na minha vida. Gostaria de agradecer a todos que, de uma forma ou de outra, colaboraram para que este trabalho se tornasse possível.*

*Em especial, ao Professor Doutor Paulo Oliveira pela forma como me orientou, pela sua disponibilidade, pelas opiniões e críticas e pela sabedoria que me transmitiu.*

*Aos meus pais, às minhas irmãs, cunhado e sobrinho pelo apoio, incentivo e paciência.*

*Aos meus amigos, em especial à Susana Gomes, por terem estado sempre presentes quando mais precisei.*

*Ao Montepio por me ter disponibilizado os dados.*



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Formulação do Modelo</b>	<b>5</b>
2.1	Ideia básica do modelo GDM . . . . .	5
2.2	Determinação de $w_i$ . . . . .	8
2.3	Determinação do consenso de grupo final . . . . .	13
<b>3</b>	<b>Exemplo numérico</b>	<b>17</b>
<b>4</b>	<b>Aplicação a um caso concreto</b>	<b>21</b>
4.1	Dados . . . . .	21
4.2	Análise de Componentes Principais . . . . .	22
4.3	Modelo de regressão logística . . . . .	23
4.4	Resultados . . . . .	24
4.4.1	Modelo 1 . . . . .	26
4.4.2	Modelo 2 . . . . .	27
4.4.3	Modelo 3 . . . . .	28
4.4.4	Modelo 4 . . . . .	29
4.5	Considerações finais . . . . .	30
4.5.1	Por cliente . . . . .	30
4.5.2	Por modelo . . . . .	34
<b>A</b>	<b>Anexos</b>	<b>35</b>





# Capítulo 1

## Introdução

A concessão de crédito desempenha um papel fundamental no desenvolvimento de uma economia, seja como uma oportunidade para as empresas aumentarem os seus níveis de produção seja como estímulo ao consumo dos indivíduos. Consequentemente, torna-se um problema primário da decisão de negócios para as instituições, como bancos comerciais e empresas que concedem cartões de crédito. O *credit scoring* é uma das técnicas utilizadas em problemas de avaliação de risco de crédito, através da ponderação de vários fatores, como a idade, profissão, rendimentos, entre outros. Este método classifica os clientes em duas categorias: os que têm condições para honrar as obrigações do empréstimo obtido, que são considerados bons clientes, e os que não reúnem as condições para cumprir as obrigações de crédito, que são considerados maus clientes. Durante a década de 1950 foram desenvolvidas diversas metodologias de classificação de crédito, mas foi com o matemático Bill Fair e o engenheiro Earl Isaac, ao fundarem a empresa *Fair, Isaac and Company*, que se deu início à implementação das metodologias, impulsionando significativamente a indústria do crédito. A fase de grande expansão verificou-se em 1970, tanto nos Estados Unidos como na Europa. A indústria do crédito ia evoluindo à medida que a sociedade ia adquirindo novos hábitos de consumo e exigindo resposta às novas necessidades.

Assim, a ideia básica dos modelos de *credit scoring* é quantificar os fatores chave que influenciam a probabilidade de incumprimento dos clientes, permitindo a classificação dos mesmos em grupos distintos e como consequência, a decisão sobre a aceitação ou não da proposta em análise. Recorrendo a dados históricos, o *credit scoring* isola as características dos solicitantes que entraram em situação de *default* produzindo, então, uma pontuação que a instituição utiliza para classificar o candidato ao crédito em termos de risco e decidir quanto à sua concessão. Deste modo, o objetivo principal do *scoring* é melhorar o processo de seleção dos bons clientes com o intuito de reduzir futuras perdas. Esta seleção é feita com recurso a alguns métodos paramétricos e não paramétricos, tais como: regressão linear, análise discriminante,

regressão logística, árvores de decisão, redes neurais, programação linear, entre outros.

A aplicação destas metodologias permite um aumento do lucro da instituição através de maiores índices de aprovação, ou seja, a probabilidade dos clientes conseguirem pagar o crédito é elevada, o que leva a uma redução dos níveis de *default*. Assim aumenta a qualidade do serviço prestado, sendo este processo facilmente compreendido pelos seus participantes. No entanto, pode existir uma escassez de qualidade dos dados, uma vez que estes modelos são desenvolvidos com base em observações presentes nas bases de dados das instituições. Estes modelos baseiam-se no princípio de que o passado prevê o futuro. A sua implementação pode acarretar custos para a instituição ou resultar em situações danosas na concessão do crédito.

O objetivo deste estudo é analisar e avaliar os níveis de risco de crédito dos candidatos. Nesta análise surgem situações de conflito entre opiniões de especialistas. Assim será necessário encontrar um consenso de grupo que represente uma opinião comum num ambiente de decisão de grupo. Para tal forma-se um grupo com dois ou mais elementos, com o objetivo de definir a natureza do problema, generalizando possíveis soluções, avaliando o potencial das soluções ou formulando estratégias para implementar as soluções, com o objetivo de produzirem um consenso de grupo. Cada elemento deste grupo é caracterizado pelas suas próprias perceções, atitudes, motivações, entre outros. Estes elementos reconhecem a existência de um problema comum e tentam chegar a uma decisão coletiva. Este grupo é designado por *Group Decision Making* (GDM).

O GDM recorre a várias técnicas para fazer esta análise, como técnicas de inteligência artificial ou outras que sejam mais adequadas. A inteligência artificial é uma área de pesquisa da ciência da computação aplicada à utilização de métodos ou dispositivos computacionais que possuam ou multipliquem a capacidade racional do ser humano resolver problemas, pensar de forma ampla, ser inteligente, entre outros. Os agentes de inteligência artificial podem ser vistos como membros do GDM, podendo dar uma avaliação sobre um dado problema, em relação a um conjunto de critérios pré-definidos, e posteriormente os respetivos resultados. Os julgamentos dos especialistas humanos são diferentes daqueles fornecidos pelos agentes inteligentes. Estes últimos são mais objetivos, pois são pouco afetados por aspetos externos, podendo gerar julgamentos diferentes apesar de os critérios usados serem os mesmos, uma vez que os parâmetros de amostragem são variáveis e instáveis.

As instituições exigem, muitas vezes, que as decisões sejam tomadas por um grupo. Estes têm opiniões individuais difusas e distintas umas das outras na tentativa de determinar a solução ótima. Devido às imprecisões do mundo real estes decisores não são capazes de avaliar as suas opiniões com um só dado numérico e, assim sendo, têm necessidade de as representar por um intervalo. Para analisar este problema de *credit scoring* vamos basear-nos numa representação imprecisa da classificação de crédito, aproximando-nos assim da imprecisão do mundo real. Suponhamos que um decisor tem uma determinada opinião, onde para a classificar terá de atribuir um valor. Este decisor, em vez de dizer que a sua opinião corresponde o valor 50, irá dizer que a sua opinião pode variar entre 45 e 52. Estas opiniões designam-se por opiniões difusas ou *fuzzy*. De modo a lidar com a imprecisão da preferência dos decisores introduzimos um modelo GDM *fuzzy*. Este permite que os membros do grupo expressem individualmente as suas preferências difusas e que se obtenha uma ponderação dos membros do grupo.

O modelo GDM *fuzzy* segue as seguintes etapas:

1. Utilização de técnicas julgadas adequadas para analisar e avaliar os problemas de decisão sobre um conjunto de critérios;
2. Formulação de diferentes juízos, uma vez que podem existir diferentes parâmetros e amostragens;
3. Utilização de técnicas de otimização clássicas com o intuito de determinar o juízo final para a tomada de decisão.

Na primeira parte deste trabalho formularemos o modelo teórico que posteriormente será analisado. Mais à frente será apresentado um exemplo numérico. Para finalizar, vamos aplicar este modelo a um caso específico real, o qual será utilizado para concluir acerca do estudo realizado.



# Capítulo 2

## Formulação do Modelo

Nesta secção iremos formular um modelo para avaliar o risco de crédito. Para resolvermos problemas de *credit scoring* consideramos uma amostra que servirá para construir um modelo que será utilizado em dados futuros. Este modelo baseia-se numa função cujos argumentos são as variáveis recolhidas, que origina a pontuação de crédito.

### 2.1. Ideia básica do modelo GDM

Suponhamos que existem  $n$  decisores que atuam como agentes inteligentes e  $m$  parâmetros para os problemas de decisão, ou seja, cada decisor utilizará um modelo que irá depender de  $m$  variáveis.

Para um dado problema de decisão, os diferentes decisores têm diferentes julgamentos sobre um determinado conjunto de dados. Deste modo, temos:

$$X = (x_1, x_2, \dots, x_m),$$

onde  $X$  representa o conjunto dos  $m$  parâmetros fornecidos pelos clientes.

Como estamos perante descrições de classificação de crédito que envolvem algumas imprecisões, recorreremos ao processo de *fuzzification*, para agregar os diferentes julgamentos dos decisores que pretendem tomar a decisão final e fazer uso integral dos diferentes juízos. Este processo consiste na conversão de um valor numérico num grau de pertença. Será útil para descrever os julgamentos dos decisores, uma vez que estes são imprecisos logo não se podem classificar com um valor numérico, mas sim com um número *fuzzy*. A *fuzzification* baseia-se nas seguintes fases: análise do problema, definição das variáveis, definição da função de pertença e na criação dos conjuntos *fuzzy*.

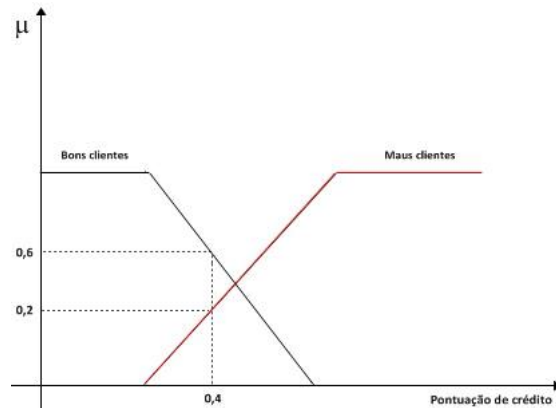
Sendo assim, introduziremos algumas definições necessárias à aplicação deste processo.

**Definição 2.1.1.** Um conjunto fuzzy,  $\tilde{A}$ , num universo  $Y$ , é caracterizado por uma função de pertença  $\mu_{\tilde{A}}(y)$  que associa a cada elemento  $y \in Y$  um número real que pertence ao intervalo  $[0, 1]$ .

O valor da função  $\mu_{\tilde{A}}(y)$  é denominado por **grau de pertença** de  $y$  em  $\tilde{A}$ .

De seguida, apresentamos um exemplo que ilustra os números fuzzy.

Suponhamos que um decisor atribuiu a um cliente uma pontuação de crédito de 0,4. Graficamente, tem-se:



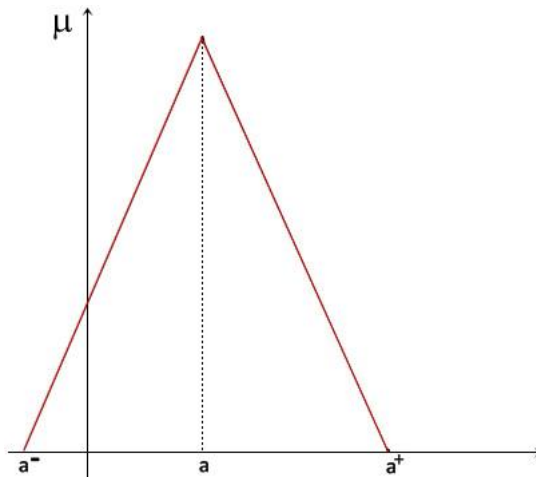
Isto significa que o cliente ao qual o decisor atribuiu uma pontuação de crédito de 0,4 é 60% bom cliente e 20% mau cliente. Caso  $\mu = 1$  então o cliente pertencia ao grupo dos bons clientes, por outro lado se  $\mu = 0$  então o cliente pertencia ao grupo dos maus clientes. Os valores intermédios indicam assim o grau de pertença do cliente ao grupo.

Existem diversas formas de representar os números fuzzy. Iremos considerar os números triangulares fuzzy.

**Definição 2.1.2.** Um número triangular fuzzy,  $\tilde{a}$ , é definido por  $(a^-, a, a^+)$ . Neste caso a função  $\mu_{\tilde{a}}(y)$  é dada por:

$$\mu_{\tilde{a}}(y) = \begin{cases} 0 & \text{se } y < a^- \\ \frac{y-a^-}{a-a^-} & \text{se } a^- \leq y \leq a \\ \frac{a^+-y}{a^+-a} & \text{se } a \leq y \leq a^+ \\ 0 & \text{se } a^+ < y \end{cases}$$

Graficamente, temos:



Deste modo, consideramos o seguinte número triangular *fuzzy*,  $\tilde{Z}_i$ , para descrever os julgamentos dos decisores, ou seja,

$$\tilde{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})$$

onde  $i$  representa o índice numérico dos decisores.

A capacidade de distinguir os bons clientes dos maus clientes é essencial para a sobrevivência e progresso destas instituições. Esta distinção faz-se recorrendo à pontuação de crédito obtida pelo *credit scoring*. Esta pontuação é uma medida da probabilidade de cumprimento/incumprimento do contrato de crédito. Se uma instituição conceder crédito a um mau cliente, ou seja, a um cliente com má pontuação de crédito, esta correrá o risco de vir a ter perdas, uma vez que o cliente pode não pagar a dívida. Por outro lado, se recusar crédito a um bom cliente, ou seja, a um cliente com boa pontuação de crédito, a instituição terá uma perda na receita futura.

Consideremos que uma instituição financeira que concede crédito fixa uma pontuação de crédito igual a 70. Suponhamos que a um determinado candidato é atribuída uma pontuação de 65. Como  $65 < 70$  este candidato será rejeitado. Por outro lado, se a um candidato for atribuída uma pontuação de 75, este será aceite pois  $75 > 70$ .

Seja  $X$  um conjunto de critérios de avaliação e  $f(x)$  a saída do agente inteligente, que é utilizada como pontuação de crédito do candidato. Consideremos  $g(X)$  as funções de classificação que são usadas para avaliar os candidatos como bons ou maus:

$$g(X) = \text{sign}(f(x) - T_\theta)$$

onde:

- $f(x)$  é o valor de saída dos decisores,
- $T_\theta$  é o limiar a partir do qual se concede o crédito.

Perante um problema de *credit scoring* o analista de crédito fixa um limiar de corte e só quando a pontuação de um candidato for superior a  $T_\theta$  é que a sua aplicação será aceite.

Sejam  $(f_1^i(X_A), f_2^i(X_A), \dots, f_k^i(X_A))$  as  $k$  pontuações de crédito diferentes do  $i$ -ésimo decisor para um candidato específico  $A$  sobre um conjunto de critérios  $X$ .

Suponhamos que um decisor atribui um número triangular *fuzzy* ao seu julgamento para um candidato específico  $A$ . De modo a usarmos todas as informações fornecidas pelo crédito, sem perda de generalidade, definimos o número triangular *fuzzy* da seguinte maneira:

$$\tilde{Z}_i = (z_{i1}, z_{i2}, z_{i3}) = \left( [\min(f_1^i(X_A), f_2^i(X_A), \dots, f_k^i(X_A))], \left[ \sum_{j=1}^k \frac{f_j^i(X_A)}{k} \right], [\max(f_1^i(X_A), f_2^i(X_A), \dots, f_k^i(X_A))] \right) \quad (2.1)$$

Neste caso, utilizamos o valor mínimo e máximo da amostra e a respetiva média da amostra.

Posteriormente iremos agregar as decisões *fuzzy* de  $n$  membros do grupo recorrendo à média ponderada, ou seja,

$$\tilde{Z} = \sum_{i=1}^n w_i \tilde{Z}_i = \left( \sum_{i=1}^n w_i z_{i1}, \sum_{i=1}^n w_i z_{i2}, \sum_{i=1}^n w_i z_{i3} \right), \quad (2.2)$$

onde  $w_i$  é o peso do  $i$ -ésimo julgamento *fuzzy*,  $i = 1, \dots, n$ .

Normalmente, os pesos satisfazem a seguinte condição de normalização:

$$\sum_{i=1}^p w_i = 1. \quad (2.3)$$

## 2.2. Determinação de $w_i$

De modo a encontrar uma só opinião que agregue as diversas opiniões individuais, queremos que a distância dessa opinião às opiniões iniciais seja mínima. Para tal iremos otimizar a média ponderada das opiniões individuais, e assim, alcançaremos a máxima concordância entre as opiniões.



**Definição 2.2.1.** Seja  $\tilde{Z} = w_i \tilde{Z}_i = (w_i z_{i1}, w_i z_{i2}, w_i z_{i3})$  e  $\tilde{Z}' = w_j \tilde{Z}_j = (w_j z_{j1}, w_j z_{j2}, w_j z_{j3})$ . A distância entre  $\tilde{Z}$  e  $\tilde{Z}'$  é dada por:

$$d_{ij} = \sqrt{\sum_{l=1}^3 (w_i z_{il} - w_j z_{jl})^2}. \quad (2.4)$$

Usando esta definição de distância, construímos o seguinte modelo de otimização, que minimiza a soma dos quadrados das distâncias entre todos os pares de julgamentos *fuzzy* com pesos:

$$\min D = \sum_{i=1}^p \sum_{j=1, j \neq i}^p d_{ij}^2 = \sum_{i=1}^p \sum_{j=1, j \neq i}^p \left[ \sum_{l=1}^3 (w_i z_{il} - w_j z_{jl})^2 \right] \quad (2.5)$$

$$s.a \sum_{i=1}^p w_i = 1 \quad (2.6)$$

$$w_i \geq 0, i = 1, 2, \dots, p. \quad (2.7)$$

Para resolver este problema de otimização não consideramos a restrição  $w_i \geq 0, i = 1, 2, \dots, p$ . Se estas variáveis forem não-negativas é satisfeita automaticamente.

Aplicando o método dos multiplicadores de Lagrange somos levados a considerar a seguinte função Langragiana:

$$L(w, \lambda) = \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[ \sum_{l=1}^3 (w_i z_{il} - w_j z_{jl})^2 \right] - 2\lambda \left( \sum_{i=1}^n w_i - 1 \right). \quad (2.8)$$

Diferenciando em ordem a  $w_i$ , definimos:

$$\begin{aligned} \frac{\partial L}{\partial w_i} &= \sum_{j=1}^n \left[ \sum_{l=1}^3 2w_i z_{il}^2 - 2w_j z_{jl} z_{il} \right] - 2\lambda \\ &= \sum_{j=1}^n \left[ \sum_{l=1}^3 2z_{il} (w_i z_{il} - w_j z_{jl}) \right] - 2\lambda \\ &= 2 \sum_{j=1}^n \left[ \sum_{l=1}^3 z_{il} (w_i z_{il} - w_j z_{jl}) \right] - 2\lambda. \end{aligned} \quad (2.9)$$

Simplificando e igualando a 0, tem-se:

$$\begin{aligned}
 \frac{\partial L}{\partial w_i} = 0 &\Leftrightarrow \sum_{j=1}^n \sum_{l=1}^3 z_{il}^2 w_i - \sum_{j=1}^n \sum_{l=1}^3 w_j z_{il} z_{jl} - \lambda = 0 \Leftrightarrow \\
 &\Leftrightarrow \sum_{j=1, j \neq i}^n \sum_{l=1}^3 z_{il}^2 w_i - \sum_{j=1, j \neq i}^n \sum_{l=1}^3 w_j z_{il} z_{jl} - \lambda = 0 \Leftrightarrow \\
 &\Leftrightarrow (n-1) \left( \sum_{l=1}^3 z_{il}^2 \right) w_i - \sum_{j=1}^n n \left( \sum_{l=1}^3 z_{il} z_{jl} \right) w_j - \lambda = 0. \tag{2.10}
 \end{aligned}$$

Consideremos as seguintes definições:

- $W = (w_1, w_2, \dots, w_n)^T$

- $I = (1, 1, \dots, 1)^T$

- 

$$b_{ij} = \begin{cases} (n-1) \left( \sum_{l=1}^3 z_{il}^2 \right) & \text{se } i = j = 1, 2, \dots, n \\ \sum_{l=1}^3 z_{il} z_{jl} & \text{se } i \neq j \end{cases}$$

Deste modo, definimos:

$$B = \begin{bmatrix} (n-1) \left( \sum_{l=1}^3 z_{1l}^2 \right) & -\sum_{l=1}^3 (z_{1l} z_{2l}) & \dots & -\sum_{l=1}^3 (z_{1l} z_{nl}) \\ -\sum_{l=1}^3 (z_{2l} z_{1l}) & (n-1) \left( \sum_{l=1}^3 z_{2l}^2 \right) & \dots & -\sum_{l=1}^3 (z_{2l} z_{nl}) \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{l=1}^3 (z_{nl} z_{1l}) & -\sum_{l=1}^3 (z_{nl} z_{2l}) & \dots & (n-1) \left( \sum_{l=1}^3 z_{nl}^2 \right) \end{bmatrix}_{n \times n}.$$

Reescrevendo a equação (2.6) ficamos com:

$$\sum_{i=1}^n w_i = 1 \Leftrightarrow w_1 + w_2 + \dots + w_n = 1 \Leftrightarrow I^T W = 1. \tag{2.11}$$

Do mesmo modo, reescrevendo a equação (2.10) tem-se:

$$\begin{aligned}
 &(n-1) \left( \sum_{l=1}^3 z_{il}^2 \right) w_i - \sum_{j=1}^n \left( \sum_{l=1}^3 z_{il} z_{jl} \right) w_j - \lambda = 0 \Leftrightarrow \\
 \Leftrightarrow &\begin{bmatrix} (n-1) \left( \sum_{l=1}^3 z_{1l}^2 \right) w_1 - \left( \sum_{l=1}^3 z_{1l} z_{2l} \right) w_2 - \dots - \left( \sum_{l=1}^3 z_{1l} z_{nl} \right) w_n - \lambda \\ - \left( \sum_{l=1}^3 z_{2l} z_{1l} \right) w_1 + (n-1) \left( \sum_{l=1}^3 z_{2l}^2 \right) w_2 - \dots - \left( \sum_{l=1}^3 z_{2l} z_{nl} \right) w_n - \lambda \\ \vdots \\ - \left( \sum_{l=1}^3 z_{nl} z_{1l} \right) w_1 - \left( \sum_{l=1}^3 z_{nl} z_{2l} \right) w_2 - \dots + (n-1) \left( \sum_{l=1}^3 z_{nl}^2 \right) w_n - \lambda \end{bmatrix} = 0 \Leftrightarrow
 \end{aligned}$$

$$\Leftrightarrow \begin{bmatrix} (n-1) \left( \sum_{l=1}^3 z_{1l}^2 \right) & -\sum_{l=1}^3 (z_{1l}z_{2l}) & \cdots & -\sum_{l=1}^3 (z_{1l}z_{nl}) \\ -\sum_{l=1}^3 (z_{2l}z_{1l}) & (n-1) \left( \sum_{l=1}^3 z_{2l}^2 \right) & \cdots & -\sum_{l=1}^3 (z_{2l}z_{nl}) \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{l=1}^3 (z_{nl}z_{1l}) & -\sum_{l=1}^3 (z_{nl}z_{2l}) & \cdots & (n-1) \left( \sum_{l=1}^3 z_{nl}^2 \right) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_n \end{bmatrix} - \lambda \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} = 0$$

$$\Leftrightarrow BW - \lambda I = 0. \quad (2.12)$$

Podemos escrever (2.6) e (2.10) na forma matricial:

$$\begin{cases} I^T W = 1 \\ BW - \lambda I = 0 \end{cases} \quad (2.13)$$

Do mesmo modo, também podemos representar (2.5) da seguinte forma:

$$D = W^T B W, \quad (2.14)$$

como podemos verificar:

$$\begin{aligned} W^T B W &= \\ &= \left[ (p-1) \left( \sum_{l=1}^3 z_{1l}^2 \right) w_1 - \left( \sum_{l=1}^3 z_{1l}z_{2l} \right) w_2 - \cdots - \left( \sum_{l=1}^3 z_{1l}z_{pl} \right) w_p \right] w_1 \\ &+ \left[ - \left( \sum_{l=1}^3 z_{2l}z_{1l} \right) w_1 + (p-1) \left( \sum_{l=1}^3 z_{2l}^2 \right) w_2 - \cdots - \left( \sum_{l=1}^3 z_{2l}z_{pl} \right) w_p \right] w_2 + \cdots + \\ &\left[ - \left( \sum_{l=1}^3 z_{pl}z_{1l} \right) w_1 - \left( \sum_{l=1}^3 z_{pl}z_{2l} \right) w_2 - \cdots + (p-1) \left( \sum_{l=1}^3 z_{pl}^2 \right) w_p \right] w_p = \\ &= (p-1) \left( \sum_{l=1}^3 z_{1l}^2 \right) w_1^2 - \left( \sum_{l=1}^3 z_{1l}z_{2l} \right) w_2 w_1 - \cdots - \left( \sum_{l=1}^3 z_{1l}z_{pl} \right) w_p w_1 - \\ &\left( \sum_{l=1}^3 z_{2l}z_{1l} \right) w_1 w_2 + (p-1) \left( \sum_{l=1}^3 z_{2l}^2 \right) w_2^2 - \cdots - \left( \sum_{l=1}^3 z_{2l}z_{pl} \right) w_p w_2 + \cdots - \\ &\left( \sum_{l=1}^3 z_{pl}z_{1l} \right) w_1 w_p - \left( \sum_{l=1}^3 z_{pl}z_{2l} \right) w_2 w_p - \cdots + (p-1) \left( \sum_{l=1}^3 z_{pl}^2 \right) w_p^2 = \\ &= (p-1) \sum_{l=1}^3 z_{1l}^2 w_1^2 - \sum_{l=1}^3 (z_{1l}z_{2l} w_2 w_1 - \cdots - z_{1l}z_{pl} w_p w_1) + \sum_{l=1}^3 (z_{2l}z_{1l} w_1 w_2 - \cdots - z_{2l}z_{pl} w_p w_2) + \end{aligned}$$

$$\begin{aligned}
 & (p-1) \sum_{l=1}^3 z_{2l}^2 w_2^2 + \dots + \sum_{l=1}^3 (z_{pl} z_{1l} w_1 w_p - z_{pl} z_{2l} w_2 w_p - \dots) + (p-1) \sum_{l=1}^3 z_{pl}^2 w_p^2 = \\
 & = (p-1) \sum_{l=1}^3 z_{1l}^2 w_1^2 - \sum_{j=1, i \neq j}^p z_{1l} z_{jl} w_j w_1 - \sum_{j=1, i \neq j}^p z_{2l} z_{jl} w_j w_2 + (p-1) \sum_{l=1}^3 z_{2l}^2 w_2^2 + \\
 & \quad \dots + - \sum_{j=1, i \neq j}^p z_{pl} z_{jl} w_j w_p + (p-1) \sum_{l=1}^3 z_{pl}^2 w_p^2 = \\
 & = (p-1) \sum_{i=1}^p \left[ \sum_{l=1}^3 z_{il}^2 w_i^2 \right] + \sum_{i=1}^p \sum_{j=1, j \neq i}^p -2z_{il} z_{jl} w_i w_j + (p-1) \sum_{j=1}^p \left[ \sum_{l=1}^3 z_{jl}^2 w_j^2 \right] = \\
 & = \sum_{i=1}^p \sum_{j=1, j \neq i}^p \left[ \sum_{l=1}^3 w_i^2 z_{il}^2 - 2z_{il} z_{jl} w_i w_j + z_{jl}^2 w_j^2 \right] = \\
 & = \sum_{i=1}^p \sum_{j=1, j \neq i}^p \left[ \sum_{l=1}^3 (w_i z_{il} - w_j z_{jl})^2 \right].
 \end{aligned}$$

Portanto,  $D = W^T B W > 0$  pelo que  $B$  é uma matriz definida positiva. Como, uma matriz definida positiva é invertível, então  $B$  é uma matriz invertível e  $B^{-1}$  também é uma matriz definida positiva.

Pela equação (2.13) tem-se:

$$BW - \lambda I = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

Multiplicando por  $B^{-1}$  e posteriormente por  $I^T$  ficamos com:

$$I^T W - \lambda I^T B^{-1} I = 0$$

Sabemos que  $I^T W = 1$ . Então:

$$\lambda I^T B^{-1} I = 1 \Leftrightarrow \lambda = \frac{1}{I^T B^{-1} I}$$

Portanto o multiplicador de Lagrange é

$$\lambda^* = (I^T B^{-1} I)^{-1}$$

E,

$$\lambda I = BW \Leftrightarrow W = \lambda B^{-1} I \Leftrightarrow W^* = \frac{B^{-1} I}{I^T B^{-1} I}.$$

Deste modo, tem-se:

$$\begin{cases} \lambda^* = (I^T B^{-1} I)^{-1} \\ W^* = \frac{B^{-1} I}{I^T B^{-1} I} \end{cases} \quad (2.15)$$

Como vimos anteriormente,  $B$  é uma matriz definida positiva e a sua inversa também o é. Neste caso, podemos garantir que o determinante de  $B$  é não-negativo. Suprimindo o mesmo número de linhas e de colunas à matriz  $B$  iremos obter uma matriz com a mesma estrutura da matriz inicial, ou seja, os elementos na sua diagonal também são elementos da matriz principal  $B$ . Deste modo, todos os menores principais serão positivos, uma vez que  $B$  é definida positiva.

**Definição 2.2.2.** Qualquer matriz  $A$  da forma  $A = sI - C$ , com  $I$  a matriz identidade  $s > 0$ ,  $C \geq 0$  e  $a_{ij}$  elementos não positivos, tal que  $s \geq \rho(C)$ , é uma M-matriz.

Em particular, se  $s > \rho(C)$  então  $A$  é uma M-matriz não singular.

Então,  $B$  é uma M-matriz não-singular, pois é uma matriz quadrada não singular com elementos fora da diagonal principal não positivos e todos os menores principais positivos [2].

**Propriedade 1.** Se  $A$  é uma matriz definida positiva, então  $A^{-1}$  existe e  $A^{-1} \geq 0$ .

Sendo assim, pelas propriedades das M-matrizes concluímos que  $B^{-1}$  é não-negativa. E, deste modo,  $W^* \geq 0$ , ou seja, todos os seus elementos são não negativos.

## 2.3. Determinação do consenso de grupo final

Para obtermos o consenso de grupo final temos de transformar os conjuntos resultantes da *fuzzification* num valor exato da pontuação de crédito. Este processo denomina-se de *defuzzification*. Existem diversas técnicas para calcular este valor e a técnica que será usada é o centróide. Nesta técnica o valor de saída é o centro de gravidade do conjunto *fuzzy*. Yager [3] propôs uma classificação que nos permitirá calcular este valor.

Suponhamos que temos  $n$  subconjuntos *fuzzy* normais convexos  $\tilde{\mu}_i$ ,  $i \in \mathbf{N} = 1, 2, \dots, n$ , isto é,

- $\forall x_1, x_2 \in X : \tilde{\mu}_i(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\tilde{\mu}_i(x_1), \tilde{\mu}_i(x_2))$  onde  $\lambda \in [0, 1]$ ,
- $\exists x_0$  tal que  $\tilde{\mu}_i(x_0) = 1$ ,

tais que:

$$\tilde{u}_i = \{z, \mu_{\tilde{u}_i}(z)\}, z \in S_i \in I. \quad (2.16)$$

Um método simples para ordenar os  $\tilde{u}_i$  consiste na definição de uma função de classificação,  $F$ , que representa cada conjunto *fuzzy*.

Seja

$$F : \mathcal{P}(I) \longrightarrow \mathbf{R}$$

onde  $\mathcal{P}(I)$  é um conjunto de subconjuntos *fuzzy* de  $I$ .  $F$  é tal que:

$$\begin{cases} F(\tilde{u}_i) < F(\tilde{u}_j) \Rightarrow \tilde{u}_i < \tilde{u}_j \\ F(\tilde{u}_i) = F(\tilde{u}_j) \Rightarrow \tilde{u}_i = \tilde{u}_j \\ F(\tilde{u}_i) > F(\tilde{u}_j) \Rightarrow \tilde{u}_i > \tilde{u}_j. \end{cases} \quad (2.17)$$

Yager propôs várias funções de classificação, onde não se assume qualquer hipótese de normalidade ou convexidade. Consideremos a seguinte função de classificação:

$$F(\tilde{u}_i) = \frac{\int_0^1 g(z) \mu_{\tilde{u}_i} dz}{\int_0^1 \mu_{\tilde{u}_i} dz} \quad (2.18)$$

onde o peso  $g(z)$  é uma medida da importância de  $z$ . Se  $g(z) = z$ , então os pesos são lineares e  $F(\tilde{u}_i)$  representa o centro de gravidade do conjunto *fuzzy*,  $\mu_{\tilde{u}_i}$ .

Um consenso do grupo *fuzzy* pode ser obtido através da equação (2.2), isto é,

$$\tilde{Z} = \sum_{i=1}^p w_i \tilde{Z}_i = \left( \sum_{i=1}^p w_i z_{i1}, \sum_{i=1}^p w_i z_{i2}, \sum_{i=1}^p w_i z_{i3} \right).$$

Para obter um valor exato da pontuação de crédito utilizamos a Classificação de Yager com  $g(x) = x$ . De acordo com esta classificação, o número triangular *fuzzy*,  $\tilde{Z} = (z_1, z_2, z_3)$ , é determinado do seguinte modo:

$$Z = \frac{\int_{z_1}^{z_3} x \mu_{\tilde{Z}}(x) dx}{\int_{z_1}^{z_3} \mu_{\tilde{Z}}(x) dx} = \frac{\int_{z_1}^{z_2} x \mu_{\tilde{Z}}(x) dx + \int_{z_2}^{z_3} x \mu_{\tilde{Z}}(x) dx}{\int_{z_1}^{z_2} \mu_{\tilde{Z}}(x) dx + \int_{z_2}^{z_3} \mu_{\tilde{Z}}(x) dx}$$

Através da definição de número triangular *fuzzy*, tem-se:

$$Z = \frac{\int_{z_1}^{z_2} \left( x \frac{x - z_1}{z_2 - z_1} \right) dx + \int_{z_2}^{z_3} \left( x \frac{z_3 - x}{z_3 - z_2} \right) dx}{\int_{z_1}^{z_2} \left( \frac{x - z_1}{z_2 - z_1} \right) dx + \int_{z_2}^{z_3} \left( \frac{z_3 - x}{z_3 - z_2} \right) dx}.$$

Resolvendo esta equação, concluímos que o consenso de grupo final é dado por:

$$Z = \frac{z_1 + z_2 + z_3}{3}. \quad (2.19)$$

Com  $z_j = \sum_{i=1}^p w_i z_{ij}$ , com  $j = 1, 2, 3$





# Capítulo 3

## Exemplo numérico

Nesta secção iremos ver uma aplicação do modelo GDM *fuzzy*. Como já referimos no Capítulo 1, as redes neuronais é uma das técnicas que pode ser usada para a análise de um problema de *credit scoring*. Deste modo, iremos considerar as redes neuronais na apresentação deste exemplo numérico.

Neste exemplo iremos considerar os seguintes algoritmos das redes neuronais:

- **BPNN** (*back-propagation neural network*)
- **RBFN** (*Radial Basis Function Network*)
- **SVMR** (*Support Vector Machine Regression*)

Para esta ilustração recorremos a estes algoritmos pois têm capacidade de obter boas aproximações. BPNN e RBFN são vistos como aproximadores universais, têm capacidade de fornecer mapeamento flexível entre entradas e saídas e dar os resultados da avaliação mais precisos do que os especialistas humanos, uma vez que os agentes inteligentes podem superar a subjetividade de especialistas em GDM.

Para cada modelo suponhamos que existem 10 parâmetros diferentes. Deste modo, criamos 30 modelos diferentes. Quando a informação de entrada de um novo candidato entra no sistema, os 30 modelos diferentes podem fornecer 30 contagens de crédito diferentes para o novo candidato. Para tal, supomos que o limiar do corte de crédito é de 60 pontos, ou seja, se a pontuação de crédito do candidato é superior a 60 o seu pedido é aceite, caso contrário o seu pedido é rejeitado. Suponhamos que as 30 pontuações de crédito, após a aplicação dos respetivos algoritmos, geradas por BPNN, RBFN e SVMR são:

$$f_{BPNN} = (57, 35; 54, 76; 59, 75; 60, 13; 59, 08; 61, 24; 56, 57; 58, 42; 60, 28; 55, 85)$$

$$f_{RBFN} = (58, 86; 60, 61; 59, 81; 57, 97; 61, 31; 62, 38; 60, 79; 59, 93; 61, 12; 61, 85)$$

$$f_{SVMR} = (59, 42; 60, 33; 58, 24; 61, 36; 63, 01; 60, 85; 62, 76; 61, 79; 63, 24; 62, 66)$$

Aplicando a equação (2.1), obtemos os seguintes números triangulares *fuzzy*, que representam as opiniões *fuzzy* dos três agentes inteligentes.

$$\tilde{Z}_{BPNN} = (z_{BPNN1}, z_{BPNN2}, z_{BPNN3}) = (54, 76; 58, 34; 61, 24)$$

$$\tilde{Z}_{RBFN} = (z_{RBFN1}, z_{RBFN2}, z_{RBFN3}) = (57, 97; 60, 46; 62, 38)$$

$$\tilde{Z}_{SVMR} = (z_{SVMR1}, z_{SVMR2}, z_{SVMR3}) = (58, 24; 61, 37; 63, 24)$$

De modo a agregarmos as três opiniões *fuzzy* num consenso de grupo recorreremos ao método de otimização, explicado na secção anterior. Para tal precisamos de calcular os respectivos pesos dos julgamentos *fuzzy*,  $W^*$ . Estes são dados por:

$$W^* = \frac{B^{-1}I}{I^T B^{-1}I}$$

Começemos por obter a matriz B, onde:

$$b_{ij} = \begin{cases} (n-1) \left( \sum_{l=1}^3 z_{il}^2 \right) & \text{se } i = j = 1, 2, \dots, n \\ \sum_{l=1}^3 z_{il} z_{jl} & \text{se } i \neq j \end{cases}$$

$$B = \begin{bmatrix} 20305 & -10522 & -10642 \\ -10522 & 21814 & -11032 \\ -10642 & -11032 & 22315 \end{bmatrix}.$$

No *software* R, calculamos a matriz inversa. De onde se obteve o seguinte:

$$B^{-1} = \begin{bmatrix} 0,3082 & 0,2973 & 0,2939 \\ 0,2973 & 0,2869 & 0,2837 \\ 0,2939 & 0,2837 & 0,2805 \end{bmatrix}.$$

Consideremos  $I = [1, 1, \dots, 1]^T$ .

Então:

$$B^{-1}I = \begin{bmatrix} 0,8994 \\ 0,8679 \\ 0,8581 \end{bmatrix}.$$

Multiplicando por  $I^T$ , ficamos com:

$$I^T B^{-1}I = 2,6254.$$

Deste modo:

$$W^* = \begin{bmatrix} 0,3426 \\ 0,3306 \\ 0,3268 \end{bmatrix}.$$

De seguida, para agregarmos as decisões *fuzzy* dos três agentes inteligentes substituímos os  $W^*$  na fórmula (2.2):

$$\tilde{Z}^* = \sum_{i=1}^n w_i \tilde{Z}_i = \left( \sum_{i=1}^n w_i z_{i1}, \sum_{i=1}^n w_i z_{i2}, \sum_{i=1}^n w_i z_{i3} \right) = (56,96; 60,03; 62,27).$$

Através do processo de *defuzzification* enunciado na secção 2.3 obtemos um valor exacto de pontuação de crédito, sob o qual tiraremos conclusões acerca da concessão de crédito ao cliente.

$$Z = \frac{z_1 + z_2 + z_3}{3} = \frac{56,96 + 60,03 + 62,27}{3} = 59,75333$$

Como a pontuação de crédito é inferior a 60 o candidato deve ser rejeitado pois é considerado um mau candidato.



## Capítulo 4

# Aplicação a um caso concreto

### 4.1. Dados

Nesta secção analisaremos uma base de dados fornecida pelo Montepio que engloba informação de 38652 clientes. Nestes dados relativos aos clientes são registados os rendimentos mensais dos trabalhadores dependentes e independentes, os montantes de responsabilidade de longo e curto prazo no Banco de Portugal, o número de elementos do agregado familiar, a antiguidade no emprego, a taxa de esforço, a idade, o LTV (*Loan To Value*), o estado civil, o regime matrimonial, situação profissional e o saldo médio de depósitos à ordem (6 meses) e de depósitos a prazo.

Para uma melhor compreensão e análise da base de dados aprofundaremos os conhecimentos de algumas variáveis em estudo. A taxa de esforço é a percentagem do rendimento familiar que é destinada ao pagamento das prestações de créditos contraídos. É calculada do seguinte modo:

$$\frac{\text{Encargos Financeiros mensais}}{\text{Rendimentos}} \times 100.$$

O *Loan to Value* é um rácio financeiro que relaciona o montante de um empréstimo com o valor do imóvel dado como garantia. Quanto maior for este valor maior será o risco.

A variável estado civil tem 4 categorias: casado, solteiro, divorciado/separado, viúvo. A variável regime matrimonial está relacionada com os indivíduos casados e tem 3 categorias: comunhão de adquiridos, separação de bens e comunhão de bens. A variável situação profissional tem 8 categorias: administrativo, operário, trabalhador por conta própria, diretor/administrativo, técnico, quadro-médio, reformado e vendedor/trabalhador serviços. Para além destes dados temos a variável dependente binária *default* que nos permite saber se estamos perante um bom ou mau cliente, ou seja, se um determinado cliente irá ou não pagar o crédito pedido. Daí que se justifique o recurso à regressão logística, objetivando a identificação dos fatores que

podem influenciar o crédito.

Numa primeira fase iremos escolher aleatoriamente 1000 clientes dos 38652 clientes, onde a primeira análise será feita com base nestes 1000 clientes. Posteriormente, devido ao excesso de variáveis, iremos fazer uma análise de componentes principais. Esta análise permite-nos reduzir o número de variáveis e eliminar a informação que não é relevante, simplificando assim a análise dos dados. De seguida recorreremos à regressão logística para estimar os coeficientes do modelo e calcular assim as probabilidades de cumprimento/incumprimento de conceder o crédito. Neste caso, definiremos os números triangulares *fuzzy* através de um intervalo de confiança de 97,5% e da respetiva estimativa, obtendo assim um limite inferior e um limite superior. Com estes números triangulares *fuzzy* iremos obter um valor exacto da pontuação de crédito, sob o qual iremos tirar conclusões acerca da concessão do crédito.

### 4.2. Análise de Componentes Principais

A análise de componentes principais (ACP) é um método que reduz o número de variáveis através da análise da base de dados original, sem que seja perdida informação relevante. Tem como objetivo utilizar a informação da base de dados e torná-la mais compacta e representativa, permitindo assim determinar as variáveis que mais influenciam a concessão do crédito. Este método consiste em transformar o conjunto das variáveis originais num conjunto de variáveis não correlacionadas, denominadas componentes principais, que derivam de combinações lineares das variáveis originais. Estas componentes são calculadas por ordem decrescente de importância, sendo que a primeira terá maior percentagem de variância e a última terá a menor percentagem. Para a aplicação deste método recorreremos ao *software R*, onde utilizamos o *package stats* versão 2.15.2. Este *package* contém funções para cálculo estatístico, sendo usado na análise estática, na modelação e simulação de dados. A função utilizada para a análise de componentes principais é a *prcomp*.

Na seguinte tabela podemos observar o resumo da análise de componentes principais.

Componentes Principais	Desvio Padrão	% de Variância	% cumulativa da variância
PC1	$4,740 \times 10^4$	0,9382	0,9382
PC2	$1,107 \times 10^4$	0,05116	0,9894
PC3	$4,456 \times 10^3$	$8,290 \times 10^{-3}$	0,9977
PC4	$2,007 \times 10^3$	$1,680 \times 10^{-3}$	0,9993
PC5	$1,246 \times 10^3$	$6,500 \times 10^{-4}$	1
PC6	22,72	0	1
PC7	18,63	0	1
PC8	9,117	0	1
PC9	7,516	0	1
PC10	1,788	0	1
PC11	0,4685	0	1
PC12	0,4122	0	1
PC13	$9,969e - 17$	0	1

A tabela sugere que as variáveis que mais influenciam o crédito são as duas primeiras, uma vez que PC1 e PC2 representam 98,94% da informação em relação às variáveis originais. No entanto, se considerarmos a terceira componente principal teremos representada 99,77% da informação inicial. Além do modelo com duas componentes principais iremos considerar três modelos com três componentes principais, uma vez que, tal como a componente principal PC3, também PC4 e PC5 representam informação inicial.

### 4.3. Modelo de regressão logística

A regressão logística é uma técnica de análise multivariada, utilizada nas situações em que a variável dependente é binária, assumindo dois resultados possíveis. O objetivo desta regressão é gerar uma função matemática que permita estabelecer a probabilidade de uma observação pertencer ao grupo previamente determinado. Possibilita a descrição do comportamento da variável independente em função das variáveis dependentes. Os coeficientes estimados por este modelo de regressão indicam a importância de cada variável, independentemente da ocorrência do acontecimento.

Assim sendo, a regressão logística consiste num modelo que relaciona um conjunto de  $p$  variáveis dependentes  $x_1, x_2, \dots, x_p$  e uma variável independente  $Y$  que assume dois casos possíveis, 0 e 1, isto é, tem apenas duas categorias (por exemplo: sim/não, aprovação/reprovação). Consideremos  $\Pi_i = P(Y = 1)$ , isto é, a probabilidade de conceder crédito a um cliente no modelo  $i$ .  $\Pi_i$  é dado por:

$$\Pi_i = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}. \quad (4.1)$$

Após a aplicação da análise de componentes principais e da regressão logística procederemos à aplicação da fórmula (2.19), ou seja,

$$Z = \frac{z_1 + z_2 + z_3}{3}.$$

Onde  $z_1$ ,  $z_2$  e  $z_3$  são o limite superior, o limite inferior e a respectiva estimativa do coeficiente, ou seja, são os valores entre os quais a probabilidade pode variar, de acordo com o decisor. Deste modo,  $Z$  dá-nos o consenso entre os três decisores.

#### 4.4. Resultados

Para fazer esta análise escolhemos quatro clientes, onde tivemos em consideração as suas diferenças e as suas semelhanças podendo compará-los e comparar as decisões geradas. Deste modo, as características destes clientes são apresentadas na seguinte tabela:

Cliente	a	b	c	d	e	f	g	h	i	j	k	l	m	n
Cliente 1	2079	0	0	3	14	41,7	38,9	31,8	0	1	1	2	374	0
Cliente 2	1629	67057	0	3	14	23,3	40,3	49	0	1	1	3	56	0
Cliente 4	5891	0	333	3	9	18,3	32,8	104,7	0	1	1	1	3931	4621
Cliente 6	2434	1972	0	3	17	27,1	46	80	0	1	1	5	9371	15054

onde:



- a - Rendimento mensal dos trabalhadores dependentes e independentes
- b - Montante Responsabilidade de longo prazo no Banco de Portugal
- c - Montante Responsabilidade de curto prazo no Banco de Portugal
- d - Número de elementos do agregado familiar
- e - Antiguidade no emprego
- f - Taxa de esforço
- g - Idade (em anos)
- h - LTV(*Loan to Value*)
- i - Default
- j - Código do estado civil
- k - Código do regime matrimonial
- l - Código da situação profissional
- m - Saldo médio de depósitos à ordem (6 meses)
- n - Saldo de depósitos a prazo

Os clientes analisados são casados, vivem em comunhão de adquiridos e têm um filho. Os rendimentos, os montantes de reponsabilidades, a taxa de esforço, o LTV e os saldos médios a prazo e à ordem dizem respeito ao casal. Os restantes dados são respeitantes ao titular da conta. O titular da conta do cliente 1 tem 38 anos, é operário e trabalha há 14 anos na mesma empresa. Possuem rendimentos médios e nao têm dívidas ao Banco de Portugal. A sua taxa de esforço é de 41,7% logo, é esta a percentagem do rendimento familiar destinada ao pagamento de créditos contraídos. Por outro lado, o LTV não é muito elevado sendo de 31,8%. O saldo médio de depósitos à ordem é de 374 euros.

O titular do cliente 2 tem 40 anos, é trabalhador por conta própria há 14 anos. Como casal têm elevadas dívidas ao Banco de Portugal apesar de possuir baixo rendimento. 23,3% do rendimento familiar é destinado ao pagamento de créditos. O LTV é de 49%, o que não revela risco elevado. O saldo médio de depósitos à ordem é de 56 euros, e não tem depósitos a prazo.

O titular do cliente 3 tem 32 anos, é administrativo e trabalha há 9 anos na mesma empresa. Os seus rendimentos são elevados e possuem dívidas ao Banco de Portugal a curto prazo. Apesar disso apenas 18,3% dos rendimentos são destinados ao pagamento de créditos. Por outro lado, o LTV é de 104%, correndo assim a instituição bancária um grande de risco uma vez que o valor pedido de empréstimo é superior ao valor de garantia. O saldo médio de depósitos à ordem é de 3831 euros e de depósitos a prazo é de 4621 euros.

O titular do cliente 4 tem 46 anos, é técnico e trabalha há 17 anos na mesma empresa. Os seus rendimentos são médios e têm dívidas a longo prazo ao Banco de Portugal. A taxa de esforço é de 27,1%. O seu LTV é de 80%, que corresponde ao limite normalmente imposto pelo banco. O saldo médio de depósitos à ordem é de 9731 euros e de depósitos a prazo é de 15054 euros.

Após uma primeira e breve análise, seria de esperar que fosse concedido crédito ao cliente 1, cliente 3 e cliente 4. Por outro lado, seria recusado o crédito ao cliente 2, uma vez que tem um elevado montante de responsabilidade ao Banco de Portugal face aos seus rendimentos mensais.

### 4.4.1. Modelo 1

Neste caso consideraremos apenas as duas primeiras componentes principais, representando estas 98,936% da informação das variáveis originais. Para calcular a probabilidade de conceder crédito a um cliente estimamos os coeficientes  $\beta$  e substituímos na fórmula (4.1). A estimação destes coeficientes é realizada com recurso à função *glm*, do package *stats* do software R. Deste modo, a probabilidade de conceder crédito a um cliente é dada por:

$$\frac{\exp(0,082 + 1,925 \times 10^{-7}x_1 + 1,41 \times 10^{-5}x_2)}{1 + \exp(0,082 + 1,925 \times 10^{-7}x_1 + 1,41 \times 10^{-5}x_2)}$$

De seguida, com recurso à função *confint* do *package stats* calculamos o intervalo de confiança de 97,5%, onde obtivemos os seguintes resultados:

	2,5%	97,5%
$\beta_0$	0,06501789	0,09898211
$\beta_1$	$-1,66014 \times 10^{-7}$	$5,5103 \times 10^{-7}$
$\beta_2$	$1,19974 \times 10^{-6}$	$2,70044 \times 10^{-5}$

Considerando as características dos clientes, e substituindo os  $x_i$ , com  $i = 1, 2$  pelo valor das componentes principais obtendo assim o valor da probabilidade de lhes ser concedido o crédito.

<b>Clientes</b>	<b>Estimativa</b>	<b>Limite inferior</b>	<b>Limite superior</b>
<b>Cliente1</b>	0,526304	0,51596	0,536627
<b>Cliente2</b>	0,517977	0,518304	0,51765
<b>Cliente3</b>	0,512898	0,514813	0,510981
<b>Cliente4</b>	0,524765	0,515918	0,533599

Aplicando a fórmula (2.19) obtivemos o seguinte:

<b>Clientes</b>	<b>Valor Final</b>
<b>Cliente1</b>	0,526297
<b>Cliente2</b>	0,517977
<b>Cliente3</b>	0,512898
<b>Cliente4</b>	0,524761

#### 4.4.2. Modelo 2

Neste caso consideraremos apenas as três primeiras componentes principais, que representam 99,765% da informação inicial. A probabilidade de conceder crédito a um cliente é dada por:

$$\frac{\exp(0,082 + 1,92 \times 10^{-7}x_1 - 1,77 \times 10^{-5}x_2 - 1,60 \times 10^{-5}x_3)}{1 + \exp(0,082 + 1,92 \times 10^{-7}x_1 - 1,77 \times 10^{-5}x_2 - 1,60 \times 10^{-5}x_3)}$$

## Capítulo 4 Aplicação a um caso concreto

---

Um intervalo de confiança de 97,5% para os coeficientes estimados é apresentado na seguinte tabela:

	2,5%	97,5%
$\beta_0$	0,06516741	0,09883805
$\beta_1$	$-1,63 \times 10^{-7}$	$5,48 \times 10^{-7}$
$\beta_2$	$-2,61 \times 10^{-5}$	$-4,33 \times 10^{-6}$
$\beta_2$	$-2,88 \times 10^{-5}$	$-3,20 \times 10^{-6}$

Considerando as características dos clientes calculamos as probabilidades de lhes ser concedido o crédito,

Clientes	Estimativa	Limite inferior	Limite superior
<b>Cliente1</b>	0,515088	0,50642	0,523746
<b>Cliente2</b>	0,518249	0,518539	0,0,51796
<b>Cliente3</b>	0,531933	0,531231	0,532638
<b>Cliente4</b>	0,516669	0,508927	0,524403

Aplicando a fórmula (2.19) obtivemos o seguinte:

Clientes	Valor Final
<b>Cliente1</b>	0,515084
<b>Cliente2</b>	0,518249
<b>Cliente3</b>	0,531934
<b>Cliente4</b>	0,516666

### 4.4.3. Modelo 3

Neste caso consideraremos apenas as duas primeiras componentes principais juntamente com a quarta componente principal, representando 99,104% da informação inicial. A probabilidade de conceder crédito a um cliente é dada por:

$$\frac{\exp(0,082 - 1,93 \times 10^{-7}x_1 - 1,41 \times 10^{-5}x_2 - 0,01995x_3)}{1 + \exp(0,082 + 1,93 \times 10^{-7}x_1 - 1,41 \times 10^{-5}x_2 - 0,0199x_3)}$$

Um intervalo de confiança de 97,5% é apresentado na seguinte tabela:

	2,5%	97,5%
$\beta_0$	0,06501972	0,09898028
$\beta_1$	$-5,51 \times 10^{-7}$	$1,66 \times 10^{-7}$
$\beta_2$	$1,20 \times 10^{-6}$	$2,70 \times 10^{-5}$
$\beta_2$	-0,05543179	0,01553522

Considerando as características dos clientes calculamos as probabilidades de lhes ser concedido o crédito,

<b>Clientes</b>	<b>Estimativa</b>	<b>Limite inferior</b>	<b>Limite superior</b>
<b>Cliente1</b>	0,52465	0,514304	0,534976
<b>Cliente2</b>	0,522764	0,523091	0,522438
<b>Cliente3</b>	0,511231	0,513146	0,509313
<b>Cliente4</b>	0,5233	0,514451	0,532135

Aplicando a fórmula (2.19) obtivemos o seguinte:

<b>Clientes</b>	<b>Valor Final</b>
<b>Cliente1</b>	0,524644
<b>Cliente2</b>	0,522764
<b>Cliente3</b>	0,51123
<b>Cliente4</b>	0,523295

#### 4.4.4. Modelo 4

Neste caso consideraremos apenas as duas primeiras componentes principais e a quinta componente principal, representando 99,001% da informação das variáveis iniciais.

A probabilidade de conceder crédito a um cliente é dada por:

$$\frac{\exp(0,082 - 1,93 \times 10^{-7}x_1 + 1,41 \times 10^{-5}x_2 - 2,71 \times 10^{-4}x_3)}{1 + \exp(0,082 - 1,93 \times 10^{-7}x_1 + 1,41 \times 10^{-5}x_2 - 2,71 \times 10^{-4}x_3)}.$$

Na seguinte tabela apresenta-se um intervalo de confiança de 97,5% para o modelo estudado:

	2,5%	97,5%
$\beta_0$	0,06501348	0,09898652
$\beta_1$	$-5,51 \times 10^{-7}$	$1,66 \times 10^{-7}$
$\beta_2$	$1,20 \times 10^{-6}$	$2,70 \times 10^{-5}$
$\beta_2$	$-1,04 \times 10^{-3}$	$4,94 \times 10^{-4}$

A probabilidade de lhes ser concedido o crédito, considerando as características dos clientes e o intervalo de confiança é dada por:

Clientes	Estimativa	Limite inferior	Limite superior
Cliente1	0,524766	0,514745	0,534769
Cliente2	0,522768	0,523106	0,522438
Cliente3	0,511244	0,513196	0,509289
Cliente4	0,523167	0,518939	0,53238

Com a aplicação da fórmula (2.19) obtivemos o seguinte resultado:

Clientes	Valor Final
Cliente1	0,52476
Cliente2	0,522771
Cliente3	0,511243
Cliente4	0,523162

## 4.5. Considerações finais

Nesta secção iremos comparar os resultados obtidos por cliente e por modelo.

Para realização dos gráficos recorreremos ao *Microsoft Excel*.

### 4.5.1. Por cliente

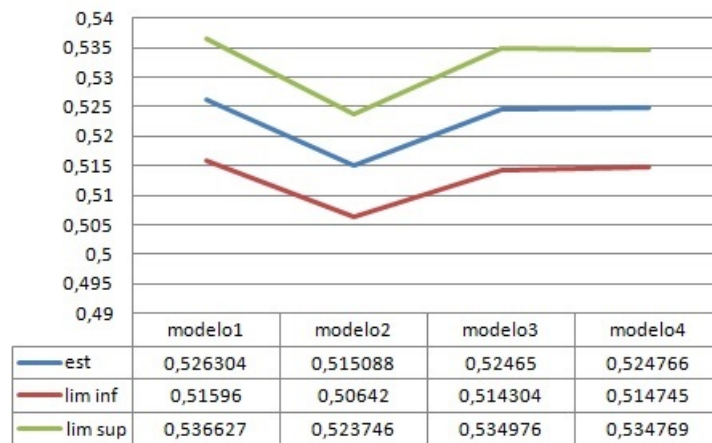
Na análise cliente a cliente iremos comparar os resultados finais obtidos nos vários modelos estudados e compararemos também as probabilidades obtidas em cada modelo.

**Cliente 1**

Começaremos por analisar os valores finais obtidos nos diversos modelos. Como podemos ver na seguinte tabela:

Cliente	Modelo1	Modelo2	Modelo3	Modelo4
<b>Cliente 1</b>	0,526297	0,515084	0,524644	0,52476

Comparando as probabilidades de ser concedido o crédito em cada Modelo, temos o seguinte gráfico:



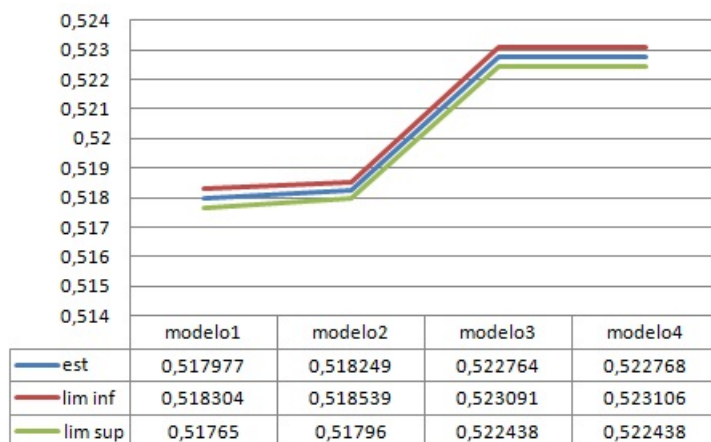
Como podemos verificar o resultado mais baixo é apresentado no Modelo 2, onde consideramos as três primeiras componentes principais. Por outro lado, o resultado mais elevado é verificado no Modelo 1, onde só consideramos as duas primeiras componentes principais, sendo os valores do Modelo 3 e Modelo 4 muito aproximados. No entanto, o cliente 1 teria mais probabilidade de lhe ser concedido o crédito se a instituição bancária considerasse o Modelo 1.

**Cliente 2**

Os valores finais obtidos em cada modelo para o cliente 2 são apresentados na seguinte tabela:

Cliente	Modelo1	Modelo2	Modelo3	Modelo4
<b>Cliente 2</b>	0,517977	0,518249	0,522764	0,522771

Comparando as probabilidades de ser concedido o crédito obtidas em cada modelo temos o seguinte gráfico:



Como podemos verificar o valor final mais baixo é apresentado no Modelo 1, quando consideramos as duas primeiras componentes principais. Por outro lado, o valor final mais elevado é apresentado no Modelo 4, quando são consideradas as duas primeiras componentes principais e a quinta componente principal. No entanto se considerassemos o Modelo 3 e o Modelo 4 a probabilidade de o crédito ser concedido ao cliente seria maior.

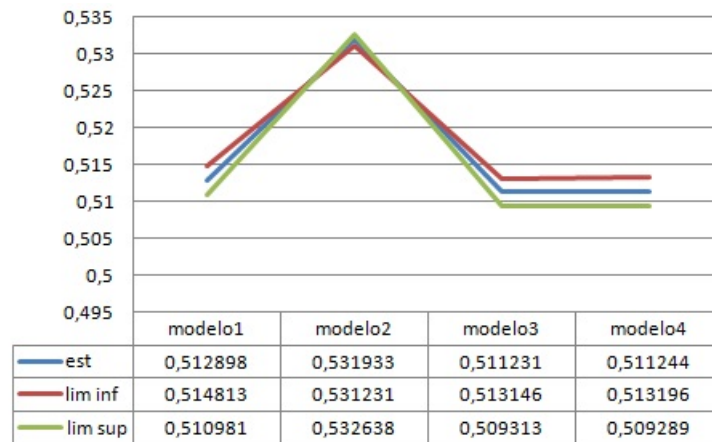
### Cliente 3

Os valores finais obtidos em cada modelo para o cliente 3 são apresentados na seguinte tabela:

Cliente	Modelo1	Modelo2	Modelo3	Modelo4
Cliente 3	0,512898	0,531934	0,51123	0,511243



No seguinte gráfico podemos observar o comportamento das probabilidades de conceder crédito a um cliente em cada modelo estudado.



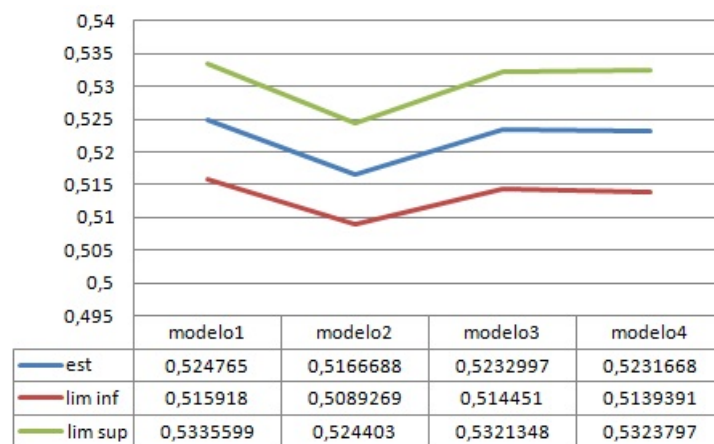
Para o cliente 3 os valores do Modelo 1 , Modelo 3 e Modelo 4 são muito próximos, no entanto o valor mais elevado é obtido no Modelo 2 e o valor mais baixo é obtido no Modelo 4. Se a instituição bancária optar pelo Modelo 2, a probabilidade de ser concedido crédito ao Cliente 3 será maior.

#### Cliente 4

Os valores finais obtidos em cada modelo para o cliente 4 são apresentados na seguinte tabela:

Cliente	Modelo1	Modelo2	Modelo3	Modelo4
Cliente 4	0,524761	0,516666	0,523295	0,523162

As probabilidades de ser concedido crédito ao cliente 4 são apresentadas no seguinte gráfico:



O cliente 4 terá maior probabilidade de lhe ser concedido o crédito no Modelo 1, tendo sido neste modelo registado o valor mais elevado. Por outro lado, o valor mais

baixo foi apresentado no Modelo2, onde se consideraram as três primeiras componentes principais.

Em geral os resultados obtidos foram muito semelhantes, sendo o Modelo 3 e o Modelo 4 os modelos que se aproximam mais.

#### 4.5.2. Por modelo

Na seguinte tabela estão apresentados valores finais para cada cliente nos respetivos modelos:

Cliente	Modelo1	Modelo2	Modelo3	Modelo4
Cliente1	0,526297	0,515084	0,524644	0,52476
Cliente2	0,517977	0,518249	0,522764	0,522771
Cliente3	0,512898	0,531934	0,51123	0,511243
Cliente4	0,524761	0,516666	0,523295	0,523162

Como podemos verificar no Modelo 1, Modelo 3 e Modelo 4 o cliente que tem maior e menor probabilidade é comum aos três modelos, cliente 1 e cliente 3, respetivamente. No entanto, no Modelo 2 o cliente 1 é o que têm menor probabilidade e o cliente 3 é o que têm maior probabilidade. Daqui podemos concluir, que o Modelo 1, Modelo 3 e Modelo 4 são modelos muito próximos sendo a sua utilização indiferente para o cliente. Por outro lado, o Modelo 2 não se aproxima destes modelos sendo os seus resultados opostos aos dos outros modelos. Como seria de esperar o cliente 1 é o cliente mais favorável à concessão de crédito, apesar de não ter rendimentos elevados, pois não apresenta quaisquer dívidas a Banco de Portugal. No entanto, o cliente 3 têm rendimentos elevados e um pequeno montante de dívidas a curto prazo ao Banco de Portugal, mas o facto de terem um LTV de 104% pode levar a que a probabilidade de conceder crédito seja menor, uma vez que o risco é maior.

# Apêndice A

## Anexos

Aqui deixamos para consulta o código feito para a escolher aleatoriamente os 1000 clientes dos 38652 clientes fornecidos pelo Montepio.

```
x<-read.csv("dados.csv",header=T,sep=";")
amostra<-array(0,c(1000,14))
ll<-sample(38652,1000,replace=T)
for (i in 1:1000)
for(j in 1:14)
amostra[i,j]<-x[ll[i],j]
```



# Bibliografia

- [1] Lean Yu, Shouyang Wang, Kin Keung Lai. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. *European Journal of Operational Research* 195 (2009) 942-959.
- [2] Berman, A., Plemmons, R.J..1979. *Nonnegative Matrices in the Mathematical Sciences*. Academic, New York.
- [3] Bortolan G., Degani, R., 1985. A review of some methods for ranking fuzzy subsets. *Fuzzy sets and Systems* 15, 1-19.
- [4] Lee, H.S., Chiu, C.C., Lu, C.J., Chen, I.F., 2002. Credit scoring using the hybrid neural discriminant technique.
- [5] Chen, M.C., Huang, S.H., 2003. Credit Scoring and rejected instances reassigning through evolutionary computation techniques.
- [6] Delgado, M. Herrera, F., Herrera-Viedma, E., Martinez, L., 1998. Combining numerical and linguistic information in group decision making. *Informatio Sciences* 107, 177-194.
- [7] Cholewa, W., 1985. Agregation of fuzzy opinions- an axiomatic approach. *Fuzzy Sets and Systems* 17, 249-258
- [8] Zhang, G., Lu, J., 2003. An integrated group decision-making method dealing with fuzzy preferences for alternatives and individual judgments for selection criteria. *Group Decision and Negotiation* 12, 501-515.
- [9] Ludmila I. Kuncheva. *Fuzzy Classifier Design*. New York: Physica-Verl., 2000.
- [10] Elizabeth Reis. *Estatística Multivariada Aplicada*. Lisboa, Novembro de 1997.
- [11] J. O. Ramsay, B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Second Edition. 2005