

UNIVERSIDADE FEDERAL DO PARANÁ  
SETOR DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE MATEMÁTICA

Thayse Duarte de Oliveira

**ANÁLISES PREDITIVAS PARA UM PROBLEMA DE  
SECURITIZAÇÃO**

Curitiba

2018

UNIVERSIDADE FEDERAL DO PARANÁ  
SETOR DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE MATEMÁTICA

Thayse Duarte de Oliveira

**ANÁLISES PREDITIVAS PARA UM PROBLEMA DE  
SECURITIZAÇÃO**

Monografia apresentada como  
requisito parcial à conclusão do  
curso de graduação em Ma-  
temática Industrial, pela Univer-  
sidade Federal do Paraná.

Orientador: Prof. Dr. Luiz Car-  
los Matioli.

Curitiba

2018

*"Seja a mudança que você quer  
ver no mundo."*

Mahatma Gandhi

# Resumo

Neste trabalho abordaremos técnicas matemáticas e estatísticas para analisar um portfólio de securitização. Securitização é uma tecnologia financeira de captação de recursos de antecipação de recebíveis.

A partir disso, é necessário estimar a variação da carteira e avaliar o comportamento dos clientes.

No primeiro caso, utilizaremos a técnica chamada Valor em Risco (VaR), que utiliza dados históricos e cenários diversos para avaliar qual é a variação mensal que a carteira pode ter em relação aos valores recebidos e esperados.

No segundo caso, calculemos a probabilidade de cada cliente entrar em inadimplência. Para tal modelo, utilizaremos uma técnica de regressão e outra de classificação. A técnica de regressão é conhecida como Regressão Logística e consiste em otimizar a função logística e assim, separar clientes bons e maus. A técnica de classificação chamada Máquina de Vetores de Suporte (SVM) cria um hiperplano que separa os clientes bons dos maus, maximizando a distância entre os dois. O SVM possui várias segmentações para cada tipo de dados.

Os dados foram cedidos pela empresa *Logos Companhia Securitizadora* e foram classificados em duas partes:

- **Desenvolvimento:** Amostra de dados utilizada para treinar os modelos.
- **Validação:** Amostra de dados utilizada para verificar se o modelo funciona.

Em nossos estudos, a Regressão Logística apresentou uma superioridade em relação ao SVM para resolver esse problema.

**Palavras-chaves:** *Aprendizado de Máquina, Regressão Logística, SVM, Máquina de Vetores de Suporte, Valor em Risco, Securitização, Análise de Risco, Otimização de Carteiras.*

# Abstract

In this work we will use mathematical and statistical techniques to analyze a securitization portfolio. Securitization is a financial technology used to raise capital through the assignment of receivables.

Given this context, it is necessary to estimate the variation of the portfolio and to evaluate the behavior of the clients.

In the first case, we will use the technique called Value at Risk (VaR), which uses historical data and several scenarios to evaluate the monthly variation that the portfolio can have in relation to the values received and expected.

In the second case, we will calculate the probability of default for each customer. For this model, we will use a regression technique and a classification technique. The regression technique is known as Logistic Regression and consists of optimizing the logistics function to separate good and bad customers. The classification technique called Support Vector Machine (SVM) creates a hyperplane that separates good customers from bad ones, maximizing the distance between the two. SVM has multiple segmentations for each data type.

The data were provided by the company *Logos Companhia Securitizadora* and were classified into two parts:

- **Development:** Sample of data used to train models.
- **Validation:** Sample of data used to verify if the model works.

In our studies, Logistic Regression presented a superiority in relation to SVM to solve this problem.

**Keywords:** *Machine Learning, Logistic Regression, SVM, Support Vector Machine, Value at Risk, Securitization, Risk Analysis, Portfolio Optimization*

# Lista de Figuras

1.1	Processo de Securitização [3]	3
2.1	Distribuição Normal [15]	8
2.2	Regressão Logística [14]	11
2.3	Máquina de Vetores de Suporte [16]	12
2.4	Truque de Kernel [16]	13
2.5	KS [17]	16

# Lista de Tabelas

2.1	Tipos de Aprendizado . . . . .	9
2.2	Matriz Confusão . . . . .	14
3.1	Seleção de Variáveis . . . . .	22
3.2	Peso das Variáveis - Stepwise . . . . .	23
3.3	Resultados Stepwise - Regressão Logística . . . . .	23
3.4	Peso das Variáveis - Correlação . . . . .	23
3.5	Resultados Correlação - Regressão Logística . . . . .	24
3.6	Resultados Correlação - SVM com Kernel Gaussiano . . . . .	24
3.7	Resultados Correlação - SVM com Kernel Linear . . . . .	25
3.8	Resultados Stepwise - SVM com Kernel Gaussiano . . . . .	25
3.9	Resultados Stepwise - SVM com Kernel Linear . . . . .	25

# Sumário

<b>Introdução</b>	<b>1</b>
<b>1 Securitização</b>	<b>2</b>
1.1 O que é . . . . .	2
1.2 Operações . . . . .	3
1.3 Elegibilidade . . . . .	4
1.4 Curva de amortização . . . . .	4
1.5 Cálculo do saldo devedor com atualização . . . . .	5
<b>2 Descrição dos métodos utilizados</b>	<b>7</b>
2.1 Revisão estatística . . . . .	7
2.2 Aprendizado de máquina . . . . .	9
2.3 Regressão stepwise . . . . .	9
2.4 Regressão logística . . . . .	10
2.5 SVM . . . . .	11
2.5.1 C-SVM . . . . .	12
2.5.2 Truque de kernel . . . . .	13
2.6 Métodos de avaliação dos modelos . . . . .	14
2.6.1 Matriz confusão . . . . .	14
2.6.2 Acurácia e erro . . . . .	14
2.6.3 Precisão . . . . .	15
2.6.4 Especificidade . . . . .	15
2.6.5 Sensibilidade . . . . .	15
2.6.6 KS . . . . .	15
2.7 VaR . . . . .	16
2.7.1 Histórico . . . . .	17
2.7.2 Paramétrico . . . . .	17
<b>3 Experimentos numéricos</b>	<b>18</b>
3.1 O problema . . . . .	18



3.2	Dados . . . . .	19
3.3	Variável de performance . . . . .	20
3.4	Seleção de variáveis . . . . .	21
3.5	Regressão logística . . . . .	23
3.6	SVM . . . . .	24
3.7	VaR . . . . .	25
<b>4</b>	<b>Conclusões e trabalhos futuros</b>	<b>27</b>

# Introdução

Risco pode ser definido como a possibilidade de perda de capital monetário. Considerando o mercado brasileiro nos dias de hoje, é necessário analisar vários cenários antes de fazer um investimento. Nas análises realizadas neste trabalho, nos encontramos do outro lado, sendo os emissores de títulos de investimento. Por isso, trabalhamos com a empresa *Logos Companhia Securitizadora* para auxiliá-los na estimativa de risco que pode acarretar em perdas e causar prejuízo aos seus investidores. Do dicionário, securitização significa ato de tornar uma dívida qualquer com determinado credor em dívida com compradores de títulos no mesmo valor. Em outras palavras, é a conversão de empréstimo e outros ativos em títulos de investimento a serem vendidas no mercado de capitais. O processo de securitização pode ser realizado com qualquer tipo de recebível, porém, neste trabalho só abordaremos recebíveis imobiliários.

Além de emitir investimentos lastreados em recebíveis imobiliários, a empresa também administra a carteira de recebíveis. Com isso, é necessário entender que para a securitizadora, o risco se aplica quando os clientes que financiaram imóveis não pagam suas dívidas, de forma que a carteira não gera recursos suficientes para o pagamento dos investidores.

Este trabalho tem como propósito estudar métodos para estimar o risco de uma das carteiras da *Logos Companhia Securitizadora*. Como um estudo inicial, trabalhamos para estimar a perda em um curto período de tempo. Para tal, escolhemos técnicas utilizadas com frequência em estudos para avaliação de possibilidade de ocorrência de perdas.

Sendo assim, no próximo capítulo apresentaremos as principais definições relacionadas ao conceito de securitização.

# Capítulo 1

## Securitização

### 1.1 O que é

Securitização é uma ferramenta de captação de recursos que transforma um fluxo de recebíveis de médio ou longo prazo em títulos mobiliários negociáveis à vista no mercado de capitais [1]. No mercado imobiliário por exemplo, uma instituição financeira ou a própria empresa originadora, faz algum tipo de financiamento imobiliário, o qual pode ser residencial, comercial ou para construções. A securitizadora adquire os recebíveis, emite os títulos de crédito no mercado financeiro conhecidos como Certificado de Recebíveis Imobiliários (CRI).

Os recebíveis possuem correções monetárias que ocorrem ao longo do tempo, o que possibilita a securitizadora repassar integral ou parcialmente tais rendimentos para seus investidores. Somente companhias securitizadoras devidamente registradas na Comissão de Valores Mobiliários (CVM) podem emitir tais títulos.

Cada CRI é vendido por um preço unitário (PU). O valor da remuneração de juros é negociada entre a securitizadora e o investidor. Esta taxa pode ser fixa ou pode ser indexada acima de algum índice de inflação como o IPCA ou IGPM, por exemplo. Chamaremos o índice utilizado de taxa de mercado. Periodicamente ou quando o título vence, o investidor recebe de volta o valor investido. Este ativo é considerado um investimento de renda fixa.

Para deixar a explicação mais clara, iremos nos referir aos endividados como clientes e aos recebedores finais da remuneração como investidores.

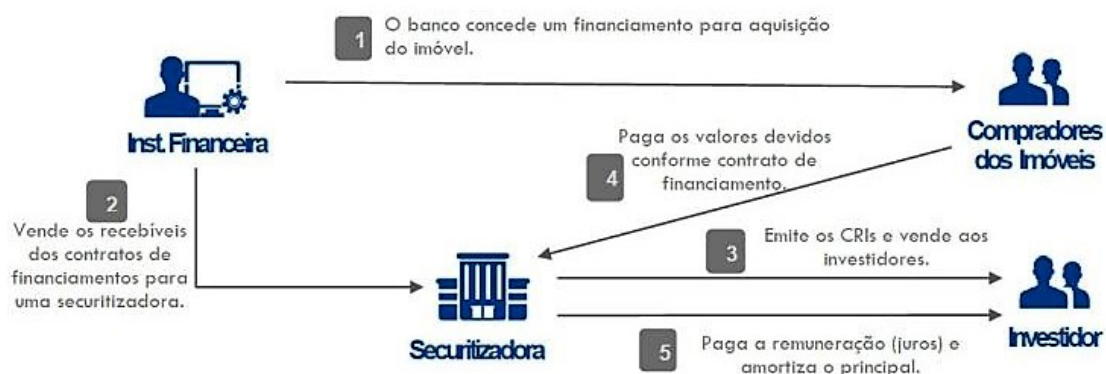


Figura 1.1: Processo de Securitização [3]

## 1.2 Operações

Cada vez que a empresa faz a securitização de recebíveis, denomina-se operação. As operações são divididas em séries que determinam as condições de retorno do investimento, sendo separadas através de um valor percentual do total investido, sendo mais comum uma série sênior e uma subordinada. A diferença entre as duas pode ser, por exemplo, a taxa de juros paga ao investidor ou o prazo de investimento.

Geralmente, a série sênior possui uma taxa de juros menor e o prazo para o vencimento do título é menor. A série subordinada apresenta uma taxa de juros remuneratórios maior, porém o prazo para a quitação da dívida com o investidor é maior. Usualmente, a quantidade de CRIs emitidos para cada série também é diferente. Normalmente, a série sênior possui mais CRIs.

Em alguns casos, os CRIs subordinados são amortizados somente após o pagamento dos CRIs seniores.

*Exemplo prático:* Sejam 10 CRIs emitidos no valor total de R\$ 10.000.000,00. As séries são divididas entre sênior e subordinada, sendo a sênior compondo 70% e a subordinada os 30% restantes dos CRIs. Neste caso, cada CRI possui o preço unitário de R\$ 1.000.000,00.

Portanto, a série sênior tem 7 CRIs e a subordinada 3. Os CRIs seniores possuem uma remuneração de 8% de juros ao ano acima do indexador IPCA. Já os CRIs subordinados possuem uma remuneração de 13,5% ao ano de juros, também acima do IPCA. Porém, o principal de R\$ 3.000.000,00 dos CRIs subordinados só começará a ser amortizada depois que o saldo devedor e os juros dos CRIs seniores forem quitados. Enquanto isso não acontece, os juros dos CRIs subordinados são pagos mensalmente.

## 1.3 Elegibilidade

Para escolher quais contratos serão securitizados, a securitizadora avalia quais clientes são ou não elegíveis. Nessa análise, alguns conceitos são aplicados para selecionar clientes que apresentam um risco menor de inadimplência. Neste caso específico, um cliente elegível é um cliente que atende as seguintes condições:

- não possui qualquer parcela vencida e não paga na data de referência.
- apresente pelo menos seis meses de histórico de pagamentos.
- apresente *Loan-to-Value* (LTV), isto é, percentual devedor em relação ao total da dívida inferior a 80%.
- apresente no máximo duas parcelas pagas com mais de 30 dias de atraso nos últimos 12 meses.
- apresente no máximo uma parcela paga com mais de 60 dias de atraso nos últimos 12 meses.
- apresente no máximo 90 dias de atraso em qualquer parcela paga nos últimos 12 meses.

Após escolher os clientes que compõem a carteira, é calculado o saldo devedor, ou seja, a soma do valor de todas as parcelas futuras de todos os clientes. Com isso, temos uma previsão de quanto a empresa securitizadora receberá mensalmente.

## 1.4 Curva de amortização

Quando a carteira elegível é definida e os juros negociados, calcula-se o valor de captação a ser repassado para a instituição financeira ou incorporadora.

Normalmente, esse valor é um percentual da somatória do valor presente da carteira de recebíveis. A partir deste valor e das informações de parcelas futuras, podemos ter uma previsão de quanto tempo demoraria para o pagamento integral dos CRI emitidos para os investidores.

Os pagamentos dos investidores são calculados considerando os valores recebidos dos clientes no mês e as despesas de cada operação. A amortização é o valor que será debitado do saldo devedor, ou seja, o pagamento menos os juros. Então é calculado o percentual de amortização mensal, que é a divisão entre o pagamento menos os juros e o saldo devedor anterior. Com esse percentual, chega-se na estimativa do prazo final do pagamento dos CRIs, em outras palavras, calcula-se quanto tempo demorará para pagar os juros e amortizar o principal dos CRIs.

*Exemplo prático:* Continuando o exemplo da sessão *Operações*, precisamos entender que os CRIs têm um saldo devedor inicial de R\$ 10.000.000,00.

Seja um cliente com uma dívida de R\$ 300.000,00 que pagará em 60 vezes. Sem considerar o juros desse financiamento e a taxa de mercado, o valor da parcela é de R\$ 5.000,00. É com a soma das parcelas de cada cliente que o pagamento ao investidor será realizado. Assim, quando os juros de cada série e as despesas de cada operação são descontados do valor total das parcelas neste mês, este é o valor que está sendo retirado do saldo devedor, que chamamos de amortização.

Sendo assim, se no mês atual, recebemos R\$ 500.000,00 como pagamento dos clientes, os juros são calculados a partir do saldo devedor, neste exemplo a operação acabou de começar e este é o primeiro pagamento. Utilizando a fórmula de juros compostos, transformamos a taxa anual de 8% em uma taxa mensal 0,6434% para a série sênior e para a subordinada de 13,5% anual em 1,0609% mensal.

Os juros (saldo devedor multiplicado pela taxa) de cada série ficam R\$ 45.038,21 e R\$74.260,18 para sênior e subordinada, respectivamente. Considerando o valor de despesas com R\$ 20.000,00. O valor da amortização da série sênior é R\$ 360.701,61. O saldo devedor ou valor residual fica R\$ 6.639.298,39 e o saldo devedor da série subordinada permanece o mesmo, até que não haja mais valor residual da outra série. Neste caso, o pagamento recebido dos clientes pode variar devido a ocorrência de inadimplência, antecipação de parcelas ou quitação da dívida.

## 1.5 Cálculo do saldo devedor com atualização

O cálculo do valor nominal unitário dos CRIs com atualização monetária é realizado diariamente considerando a taxa de mercado, taxa de amortização e taxa de juros de rendimento do investimento. O dia de pagamento de cada operação é chamado data de aniversário.

O valor diário de cada CRI é calculado corrigindo o valor do CRI da última data de aniversário pela taxa de mercado e somando os juros calculados em cima do valor corrigido. Todas as taxas são convertidas de taxas anuais para mensais e mensais para diárias. O índice utilizado para atualizar a taxa diária é a contagem de dias desde a última data de aniversário até a data atual dividido pela quantidade de dias no mês atual. Caso não haja data de aniversário anterior, é utilizada a data de integralização. Esta data é o dia em que a empresa securitizadora recebe o valor dos CRIs adquiridos pelo investidor.

O índice inflacionário que compõe a taxa de mercado possui defasagem de dois meses, isto é, se estamos no mês de janeiro, utilizaremos a taxa de mercado de

novembro. Caso ainda não tenha sido divulgada, é utilizada a taxa anterior.

Em meses em que o valor de recebíveis é maior do que o previsto, um adicional pode ser feito ao pagamento mensal do investidor, esse adicional é denominado amortização extraordinária e é descontado do saldo devedor. A amortização ou o valor que será descontado do saldo devedor da empresa para com o investidor, é obtida multiplicando a taxa de amortização do mês de referência calculada na curva pelo valor diário do CRI. Assim, após subtrair os valores de amortização e amortização extraordinário do valor atual do CRI, temos o valor atualizado do CRI que também é chamado de residual.

# Capítulo 2

## Descrição dos métodos utilizados

Neste capítulo faremos inicialmente, a revisão de alguns conceitos básicos para então apresentar os métodos que usaremos para quantificar risco na área relacionada a securitização.

### 2.1 Revisão estatística

Antes de apresentar os resultados dos estudos, iremos relembrar e introduzir alguns princípios para o melhor entendimento das informações que serão apresentadas posteriormente. Tais conceitos podem ser encontrados em qualquer livro de estatística básica, aqui utilizamos o livro *Estatística Básica* dos autores Pedro A. Morettin e Wilton de O. Bussab da Editora Saraiva.

**Média:** A média aritmética é a soma de todos os elementos de uma sequência dividido pelo número de elementos do conjunto.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

**Variância:** A variância é usada para medir a dispersão dos dados em torno da média.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

**Desvio Padrão:** O desvio padrão é definido como a raiz quadrada positiva da variância.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$



**Correlação:** Dados  $n$  pares de valores  $(x_1, y_1) \dots (x_n, y_n)$ , chamaremos de coeficiente de correlação entre as duas variáveis  $x$  e  $y$  a

$$\rho_{xy} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma_x} \right) \left( \frac{y_i - \bar{y}}{\sigma_y} \right)$$

ou seja, a média dos produtos dos valores padronizados das variáveis.

Tal medida varia no intervalo  $[-1, 1]$  e quanto mais próximo o resultado estiver de 1, mais correlacionadas positivamente são as variáveis, da mesma maneira que os resultados perto de -1 são correlacionados negativamente.

**Probabilidade:** A probabilidade, que chamaremos de  $P_a$ , é a chance de ocorrência de um evento aleatório.

Esta chance é um número real que varia de 0 a 1, então para qualquer evento  $a$  temos que

$$0 < P_a < 1$$

Consideramos então que  $P_a = 1$  é o *evento certo* e  $P_a = 0$  é *evento impossível*.

**Distribuição Normal:** Dizemos que a variável aleatória  $x$  tem distribuição normal com parâmetros  $\bar{x} = \mu$  e  $\sigma^2$ , com  $-\infty < \mu < \infty$  e  $0 < \sigma^2 < \infty$  se a densidade é dada por

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in (-\infty, \infty)$$

A distribuição normal tem a seguinte forma:

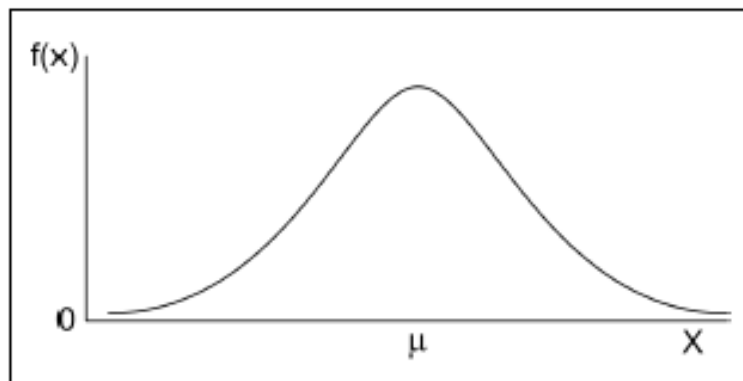


Figura 2.1: Distribuição Normal [15]

## 2.2 Aprendizado de máquina

Da mesma maneira que os seres humanos aprendem naturalmente através da experiência, o aprendizado de máquina (*Machine Learning*) apresenta este mesmo conceito. Tal técnica utiliza métodos computacionais para procurar padrões em dados e pode ser aplicada para prever riscos, processamento de imagens, manutenção preventiva de veículos e vários outros problemas cotidianos.

Os métodos são classificados em dois tipos, aprendizado supervisionado e não supervisionado. A diferença entre os dois é que no aprendizado supervisionado, o algoritmo recebe a resposta para criar o padrão, ou seja, recebe um conjunto de treinamento para identificar os padrões e sempre há uma variável resposta, como sim ou não, cliente bom ou mau e etc. Já no outro caso, o algoritmo procura padrões e agrupa dados semelhantes sem utilizar uma variável resposta para o treinamento do modelo.

Tabela 2.1: Tipos de Aprendizado

<b>Supervisionado</b>	<b>Não Supervisionado</b>
Regressão	Clusterização
Classificação	

Os métodos utilizados neste trabalho são regressão logística da classe regressão e o SVM da classe de classificação. Ambos são comumente utilizados para treinar modelos de probabilidade de inadimplência.

Não utilizaremos métodos de aprendizado não supervisionado pois estamos interessados em classificar os dados entre bons e maus pagadores no caso de inadimplência.

Segundo a revista Tecnologia de Crédito da Serasa Experian [8], o SVM e a regressão logística são consideradas como a segunda e terceira melhor técnicas para modelos de crédito, ficando atrás apenas das Redes Neurais.

## 2.3 Regressão stepwise

A regressão stepwise é comumente utilizada para seleção de variáveis. Ela consiste em adicionar e remover variáveis de um modelo de regressão linear a partir de um critério que pode ser o teste para comparação de duas variáveis, coeficiente de correlação, erro quadrático, critério de informação Bayesiano ou o critério de informação de Akaike.

É mais comum utilizar o teste para comparação de duas variáveis, que também é chamado de teste F. O teste F analisa a variância entre dois conjuntos de dados diferentes e os compara utilizando o teste de hipóteses. Neste caso, a avaliação classifica a importância da variável de acordo com a variável resposta, incluindo-a ou excluindo-a do modelo.

O modelo de regressão stepwise é localmente ótimo, ou seja, ele encontra um bom conjunto de variáveis, porém não é globalmente ótimo, o que significa que não encontra o melhor conjunto.

A função *stepwisefit* do software MATLAB fornece uma resposta binária, indicando como 1 as variáveis mais relevantes e 0 caso contrário.

## 2.4 Regressão logística

A regressão logística é uma das técnicas mais utilizadas para obter bons resultados através do aprendizado de máquina. É um método bastante simples sendo atualmente, o mais utilizado pelos bancos brasileiros.

Tal método consiste em estimar a probabilidade de ocorrência de um evento com base em um conjunto de características, ou seja, a chance de um dado pertencer a uma determinada classe.

A variável resposta  $y$  é binária e separa os dados em dois conjuntos, sendo que:

$y = 0$  significa que  $x$  pertence a “classe positiva”, e

$y = 1$  significa que  $x$  pertence a “classe negativa”.

Dessa forma, toma-se a função logística que calcula a probabilidade de um cliente ser inadimplente considerando que  $x$  ocorreu

$$P_a = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

Onde  $\beta_1, \dots, \beta_n$  representam os coeficientes das características ou variáveis  $x_1, \dots, x_n$ , com  $x = (1, x_1, \dots, x_n)$  e  $\beta_0$  é uma constante, também chamada de intercepto. A função  $\beta^T x$  é chamada de escore e quanto maior o escore, menor a probabilidade de pertencer a classe negativa. Para a função acima, prevemos  $y = 1$  se  $P_a$  for maior que 50% e  $y = 0$  caso contrário.

Para encontrar o melhor parâmetro  $\beta$ , é necessário minimizar a seguinte função:

$$f(\beta) = - \sum_{i=1}^n [y_i \log(P_a) + (1 - y_i) \log(1 - P_a)]$$

A função *mnrfit* do software MATLAB fornece tais parâmetros.

Após estimar os melhores coeficientes, é fundamental testar a eficiência do modelo em uma amostra diferente. Calcula-se a função logística para tais dados e verifica-se se a probabilidade calculada é coerente com a variável resposta.

A função logística tem a seguinte forma:

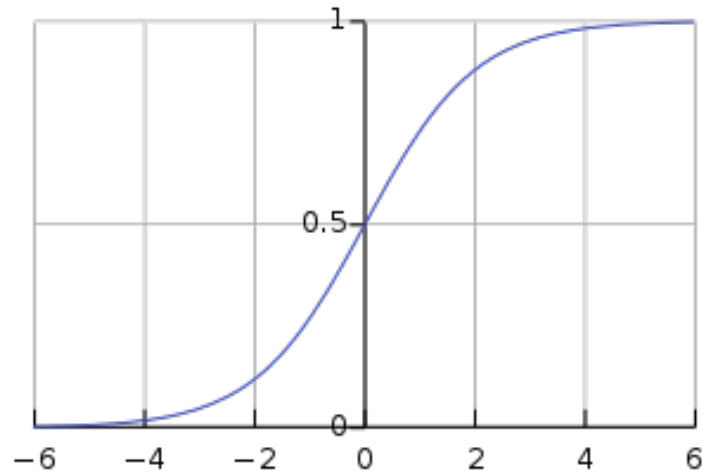


Figura 2.2: Regressão Logística [14]

## 2.5 SVM

O SVM é a técnica de classificação mais utilizada em aprendizado de máquina. O principal objetivo do método é obter um hiperplano que separa os dados entre classe positiva e negativa, assim como na regressão logística mencionada anteriormente. Casos em que a função é linearmente separável são raros e também não existe garantia de que eles possam ser separados mesmo em dimensões mais altas. Utilizaremos o C-SVM, que é um método de regularização em conjunto com o Truque de Kernel, que apresentaremos posteriormente.

O hiperplano que procuramos tem a forma:

$$w^T x + b = 0$$

Tal função pode nos ajudar a classificar novos casos de características, onde diremos que:

$x$  é da classe positiva, se  $w^T x + b > 0$  e

$x$  é da classe negativa, se  $w^T x + b < 0$

Nesse caso, a variável resposta é 1 para positivo e -1 para negativo.

É necessário acrescentar uma restrição para obter a margem máxima, isto é, a maior separação entre os dados positivos e negativos. Deste modo, a restrição dada pode ser escrita como:

$$y(w^T x + b) \geq 1 \text{ para todo } x$$

A figura a seguir fornece uma ilustração do problema

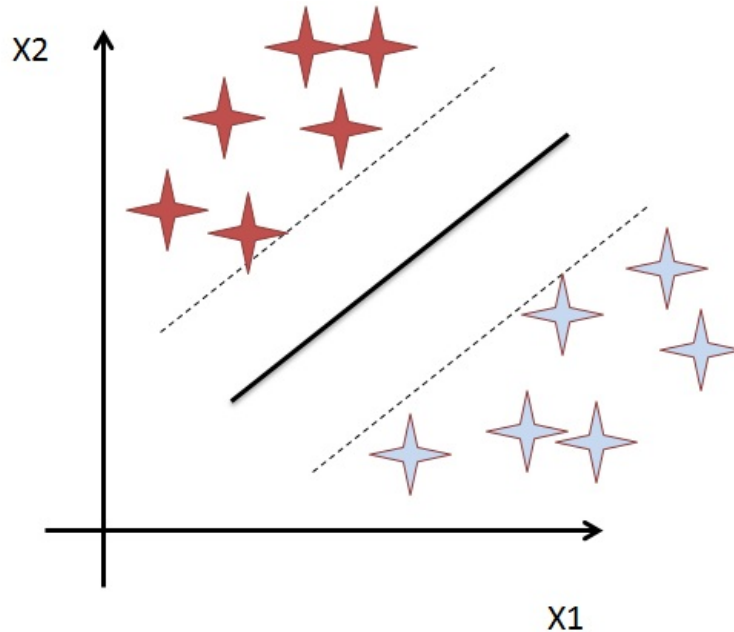


Figura 2.3: Máquina de Vetores de Suporte [16]

Neste exemplo, os dados são linearmente separáveis. As cores são as classes de dados, os clientes maus são representados pelas estrelas vermelhas e os clientes bons pelas azuis. A linha preta é o hiperplano que separa as classes. As linhas pontilhadas são a representação da maior distância entre os dois tipos de clientes.

### 2.5.1 C-SVM

Quando os dados não são separáveis, utilizamos  $m$  variáveis de folga,  $(\xi_i \geq 0)$ , para penalizar a função objetivo e dar uma folga as restrições, sendo  $C$  uma constante que multiplica todos eles. Neste caso temos o seguinte problema:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.a} \quad & y^i(w^T x + b) \geq 1 - \xi_i \end{aligned}$$

$$\xi_i \geq 0$$

$$i = 1, \dots, m$$

### 2.5.2 Truque de kernel

Quando a função não é linearmente separável, resolvemos o problema aplicando o truque de kernel. Tal truque consiste em utilizar uma transformação linear para mover o problema para outra dimensão e assim conseguir separar os dados e trazer os resultados de volta para a dimensão inicial. Como ilustrado na imagem a seguir:

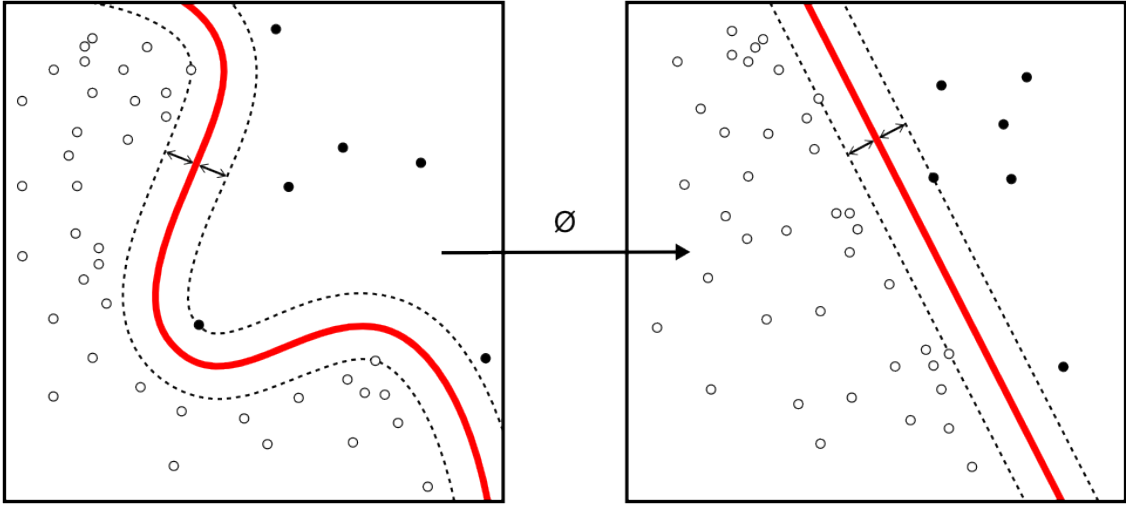


Figura 2.4: Truque de Kernel [16]

Onde  $\phi$  é o kernel utilizado na transformação deste exemplo.

Os kernels mais utilizados são o linear, polinomial e gaussiano, também chamado de função de base radial (RBF). Neste trabalho, comparamos suas performances. A função *fitcsvm* do software MATLAB fornece o modelo SVM otimizando o kernel selecionado e o parâmetro C da função objetivo.

Utilizamos os três kernels mencionados acima, os quais são dados por:

$$\text{Linear: } K(x, z) = x^T z$$

$$\text{Polinomial: } K(x, z) = (\gamma(x^T z + 1))^d$$

$$\text{RBF: } K(x, z) = e^{-\gamma \|x - z\|^2}$$

Com isso, podemos reescrever o problema da seguinte maneira:

$$\min_{\alpha} \frac{1}{2} \alpha^T H \alpha - e^T \alpha$$

$$\text{s.a } y^T \alpha = 0$$

$$0 \leq \alpha \leq C$$

com  $H(i, j) = y_i y_j K(x^i, x^j)$  e  $e_j = 1, j = 1, \dots, m$

## 2.6 Métodos de avaliação dos modelos

Nessa seção apresentaremos alguns métodos e maneiras de avaliar os modelos e assim verificar a estabilidade dos mesmos.

### 2.6.1 Matriz confusão

A matriz confusão serve para verificar quais são os valores reais em comparação com o que o modelo classificou. Com ela podemos verificar vários parâmetros que nos ajudarão a decidir se o modelo está discriminando e separando as classes corretamente.

Tabela 2.2: Matriz Confusão

Classe	Predita Positiva	Predita Negativa
Verdadeira Positiva	Vp	Fn
Verdadeira Negativa	Fp	Vn

Em que Vp é verdadeiro positivo, ou seja, a classificação positiva está correta, Fn é falso negativo, quando a amostra é classificada como negativa e é positiva. Fp é o falso positivo, onde o modelo classificou negativa e é positiva. E por último, Vn é o verdadeiro negativo, ou seja, a classificação negativa está correta.

### 2.6.2 Acurácia e erro

A acurácia também chamada de taxa de acerto é calculada da seguinte maneira:

$$ACC = \frac{Vn + Vp}{Vn + Vp + Fp + Fn}$$

O erro é calculado utilizando a fórmula da acurácia, sendo o complementar da mesma, e deverá estar próximo de zero.

$$Err = 1 - ACC$$

### 2.6.3 Precisão

Relevância da quantidade de verdadeiros positivos classificados corretamente.

$$\frac{Vp}{Vp + Fp}$$

### 2.6.4 Especificidade

Especificidade ou taxa de verdadeiros negativos mede a quantidade de negativos verdadeiros corretamente classificados.

$$\frac{Vn}{Vn + Fp}$$

### 2.6.5 Sensibilidade

Sensibilidade ou taxa de verdadeiros positivos mede a proporção de positivos verdadeiros corretamente classificados.

$$\frac{Vp}{Vp + Fn}$$

Para que o classificador seja bom, todas as taxas (excluindo o erro) precisam estar próximas de 100%.

### 2.6.6 KS

O teste Kolmogorov–Smirnov mais conhecido como KS, serve para verificar se o modelo separa corretamente a classe positiva da negativa. Porém, ele não utiliza a classificação em si e sim o score ou probabilidade. Comparando a variável resposta real, ordena-se o conjunto em forma decrescente, e são criadas 10 classes. Nestas classes, verifica-se quantos registros realmente pertencem a classe positiva e quantos pertencem a classe negativa.

$$KS_i = | \%Acumulado\ de\ Bons - \% Acumulado\ de\ Maus |$$



O KS do modelo é a máxima diferença entre o acumulado de bons e maus, isto é, o máximo  $KS_i$ .

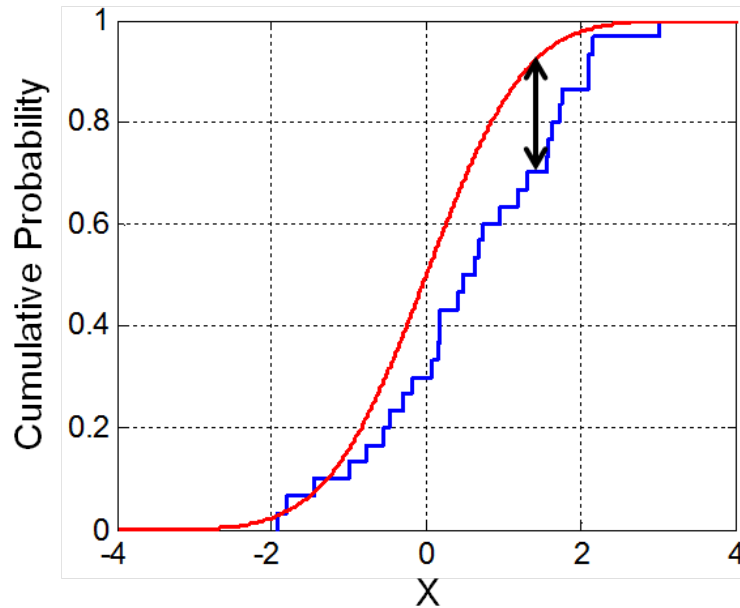


Figura 2.5: KS [17]

As curvas representam as probabilidades acumuladas de cada classe, onde a curva superior indica os bons pagadores e a inferior indica os maus. A maior separação entre elas indicada pela seta representa o KS deste exemplo.

## 2.7 VaR

O Valor em Risco também chamado de VaR, é uma técnica estatística assim como as apresentadas anteriormente, porém ela tem uma abordagem diferente. O VaR é um método para análise de risco da carteira em um determinado intervalo de tempo com um certo intervalo de confiança (um intervalo de estimativas prováveis), quando não há alterações no mercado.

*Exemplo prático:* Considere uma carteira com um VaR diário de R\$ 10.000,00 com 95% de nível de confiança. Isso significa que:

Existe uma chance/probabilidade de apenas 5% que a perda da carteira será pior que R\$ 10.000,00 no dia seguinte, ou em outras palavras, com 95% de confiança, a perda não será maior que R\$ 10.000,00 no dia seguinte.

O intervalo de confiança deve ser escolhido levando em consideração o que representa as chances da perda ser menor que o esperado (resultado dado pelo modelo). Se o intervalo de confiança é muito grande, a estimativa será, também, alta. E

se for muito pequeno, haverá muitas chances de que a estimativa dada pelo modelo não ser coerente com os dados reais.

Existem três maneiras de estimar o VaR, paramétrico, simulação histórica e simulação de Monte Carlo. Neste trabalho, não iremos calcular a simulação de Monte Carlo. Para calcular o VaR é necessário ter informações históricas de retornos ou perdas.

### 2.7.1 Histórico

O cálculo do VaR histórico é realizado ordenando-se os dados de forma decrescente e utilizando um nível de confiança escolhido.

*Exemplo prático:* Suponha que temos uma base de dados com 100 amostras e queremos calcular o VaR com 95% de confiança, ordenaremos o vetor e aplicaremos o ponto de corte no item 95. Assim, o ponto de corte será sempre no item proporcional a 95% da amostra. Caso este número não seja inteiro, arredondamos pra baixo.

Este método foi calculado utilizando o Microsoft Excel.

O método Bootstrap, semelhante ao VaR histórico, presume que os padrões do passado irão continuar a acontecer. Tal modelo utiliza cenários montados por meio de sorteio sem reposição com base nos dados históricos, e então calcula o VaR histórico para cada um deles. Este método foi implementado utilizando o MATLAB.

### 2.7.2 Paramétrico

O VaR paramétrico assume que os dados obedecem uma distribuição normal. Com isso, calcula-se a função inversa da distribuição normal, utilizando o desvio padrão, a média dos dados e a probabilidade que é dada pelo percentual complementar do intervalo de confiança.

Este método foi calculado no Microsoft Excel.

Também calculamos vários cenários do VaR paramétrico utilizando o MATLAB, tais cenários foram criados por meio da geração de valores aleatórios da distribuição normal com média e desvio padrão dos dados históricos.

# Capítulo 3

## Experimentos numéricos

### 3.1 O problema

Neste trabalho, como estudo de caso, realizamos uma parceria com a empresa *Logos Securitizadora S.A* que nos forneceu os dados para o estudo do problema. Atualmente, qualquer empresa que trabalha com mercado financeiro precisa de algum tipo de avaliação de risco. O risco também pode ser descrito como possibilidade de perder dinheiro. Verificamos que precisávamos abordar os dados de duas maneiras:

- Visão Cliente
- Visão Carteira

Na visão cliente, avaliamos qual a probabilidade de cada cliente deixar de pagar sua dívida e consequentemente, trazer prejuízo para empresa. Com essa informação, é possível avaliar o comportamento de cada cliente e ter uma prévia de quanto a empresa pode perder no mês, assim como avaliar quais clientes podem ser selecionados para realizar operações futuras. Para apresentar tais resultados, utilizamos os métodos de aprendizagem de máquina, regressão logística e o SVM.

Já na visão carteira, avaliamos as perdas mensais em um histórico de 4 anos, calculando a variação percentual entre o valor esperado e o recebido. Aplicamos o VaR para prever a variação nos meses subsequentes, criando cenários para avaliar o risco mínimo e máximo da carteira com um certo intervalo de confiança.

Apresentaremos a seguir, as informações necessárias para entender melhor como o problema foi resolvido e em seguida, seus resultados.

## 3.2 Dados

Recebemos uma base de dados com todos os clientes de um bairro planejado. Tais dados possuem 553 registros.

Todos os clientes com o saldo para quitação ou saldo recebido igual a zero foram excluídos, por terem quitado ou acabado de comprar o lote. Restando assim, 515 registros para a avaliação.

Analizamos um total de 44 variáveis, descritas abaixo:

**Empreendimento** : Contratual - Nome do condomínio.

**Bloco** : Contratual - Quadra do lote.

**Unidade** : Contratual - Número do lote.

**Área Priv.** : Contratual - Área do lote.

**Contrato** : Contratual - Número do contrato.

**Cliente** : Cadastral - Nome do cliente.

**Documento** : Cadastral - CPF do cliente.

**Dt.Nasc.** : Cadastral - Data de nascimento do cliente.

**Cônjuge** : Cadastral - Nome do cônjuge do cliente.

**Dt.Nasc.** : Cadastral - Data de nascimento do cônjuge do cliente.

**Endereço** : Cadastral - Nome da rua.

**Número** : Cadastral - Número da casa.

**Bairro** : Cadastral - Bairro atual.

**Complemento** : Cadastral - Complemento do endereço.

**Cidade** : Cadastral - Cidade em que o cliente reside.

**Estado** : Cadastral - Estado em que o cliente reside.

**Cep** : Cadastral - CEP do endereço atual.

**Telefone(s)** : Cadastral - Telefone.

**Status Contrato** : Contratual - Descrição da situação do contrato, “ativo” para contratos em dia e “inadimplente” para contratos com qualquer valor em atraso.

**Classificação** : Contratual - Descrição da situação do contrato, sendo “alienação fiduciária” a marcação de contrato securitizado.

**Data Assinatura** : Contratual - Data de início do contrato.

**Nº Meses Assinatura** : Contratual - Número de meses entre a data atual e a assinatura do contrato.

**Nº Meses Adimp.** : Contratual - Número de meses em que o cliente pagou em dia.

**Dt.Último Venc.** : Contratual - Data do vencimento da última parcela.

**Nº Meses a Vencer** : Contratual - Numero de meses entre a database e a data da última parcela.

**Valor Atraso :** Contratual - Valor total em atraso.

**Nº Parc. Atraso :** Contratual - Número de parcelas que estão em atraso.

**Total Valor Financ. :** Contratual - Valor total de financiamento.

**Data Financ. :** Contratual - Data de financiamento.

**Valor Rec Atualiz :** Contratual - Valor recebido atualizado.

**Saldo p/ Quitação :** Contratual - Saldo para quitação do lote sem atualização monetária.

**Preço Tabela :** Contratual - Preço de tabela dos lotes do condomínio.

**Valor Compra :** Contratual - Valor de compra do contrato.

**Valor Compra Atualiz :** Contratual - Soma de todas as parcelas com correção monetária.

**% Garantia :** Contratual - Percentual de garantia.

**Total Valor Futuro :** Contratual - Soma de valor futuro.

**Casado :** Construída - Marcação binária, se o cliente é casado (1) e (0) caso contrário.

**LTV :** Construída - Saldo para quitação dividido pelo valor de compra atualizado.

**Renegociação :** Construída - Marcação binária, se houve renegociação (1) e (0) caso contrário.

**Cessão :** Construída - Marcação binária, se houve cessão (1) e (0) caso contrário.

**Idade :** Construída - Idade do cliente.

**Performance :** Construída - Marcação de cliente em default ou não.

**Quantidade de meses no contrato :** Construída - Número de meses entre a assinatura do contrato e a data da última parcela.

**LTV Atualizado :** Construída - Saldo para quitação dividido pela soma de saldo para quitação e valor recebido atualizado.

Os dados foram distribuídos em dois grupos:

1. Dados para a regra de decisão, chamado de conjunto de treinamento ou desenvolvimento. Formado por 80% dos dados, 412 casos.
2. Dados para teste ou validação. Formado pelos 20% restantes, 103 casos.

### 3.3 Variável de performance

Como a variável “Status do contrato” considera qualquer tipo de atraso de qualquer valor, não poderíamos utilizá-la como variável resposta. Para construir a variável de performance, utilizamos indicadores semelhantes aos de elegibilidade, mas desconsideramos atrasos menores que 30 dias e o *Loan-to-Value*.

Assim, um cliente considerado mau satisfaz as seguintes condições:

- Qualquer parcela em aberto com mais de 30 dias de atraso.
- Possui uma ou mais parcelas pagas com mais de 90 dias de atraso nos últimos 12 meses.
- Possui duas ou mais parcelas pagas com mais de 60 dias de atraso nos últimos 12 meses.
- Possui três ou mais parcelas pagas com mais de 30 dias de atraso nos últimos 12 meses.

### 3.4 Seleção de variáveis

Como algumas dessas variáveis possuem dados pessoais, não iremos utilizá-las. Para iniciar os testes, escolhemos 23 variáveis numéricas. As variáveis “Status Contrato”, “Nº Parc. Atraso” e “Valor Rec. Atualiz.” foram retiradas da análise por serem consideradas variáveis resposta. Ou seja, se o cliente possui parcelas em atraso ou valor em atraso, intuitivamente ele é um cliente mau.

Para a selecionar as variáveis que entraram nos modelos, consideramos dois métodos. O primeiro é a regressão stepwise, a qual avalia quais variáveis são mais relevantes, utilizando o teste F. A segunda maneira é verificar quais variáveis tem uma correlação maior que 10% com a performance. As variáveis selecionadas serão marcadas como 1 nos campos **Stepwise** e **Correlação**. Ambos selecionaram 5 variáveis.

Na tabela a seguir, os resultados serão apresentados:

Tabela 3.1: Seleção de Variáveis

Número	Variável	% Correlação	Stepwise	Correlação
1	Área Priv.	-4%	0	0
2	Idade	2%	1	0
3	Casado	10%	0	1
4	Nº Meses Assinatura	-33%	1	0
5	Nº Meses Adimp.	36%	1	1
6	Nº Meses a Vencer	-7%	0	0
7	Valor Atraso	-38%	0	0
8	Nº Parc. Atraso	-44%	0	0
9	Valor Rec. Atualiz.	0%	0	0
10	Saldo p/ Quitação	-30%	0	0
11	LVT	-18%	0	0
12	1-LTV	18%	0	1
13	Preço Tabela	-6%	0	0
14	Valor Compra	20%	0	1
15	Valor Compra Atualiz.	-30%	1	0
16	% Garantia	11%	0	1
17	Total Valor Futuro	-8%	0	0
18	1- LTV Atualizado	8%	0	0
19	LVT atualizado	-8%	0	0
20	Total	-18%	0	0
21	Cessão	1%	0	0
22	Renegociação	-29%	1	0
23	Meses no contrato	-20%	0	0

### 3.5 Regressão logística

O modelo de regressão logística foi calculado utilizando a função *mrnfit* do MATLAB. Fizemos testes para os dois conjuntos de variáveis selecionados. Primeiro, apresentaremos os resultados da regressão stepwise que se saiu melhor que os outros modelos apresentados.

Tabela 3.2: Peso das Variáveis - Stepwise

Variável	Número	Beta Stepwise
Intercepto	0	-5.3331
Idade	2	-0.019116
Nº Meses Assinatura	4	0.073713
Nº Meses Adimp.	5	-0.66583
Valor Compra Atualiz.	15	0.000026389
Renegociação	22	1.1054

E os resultados dos modelos são:

Tabela 3.3: Resultados Stepwise - Regressão Logística

Acurácia	83%
Erro	17%
Sensibilidade	83%
Precisão	78%
Especificidade	88%
KS	66%

O segundo melhor modelo, foi de regressão logística com a seleção de variáveis a partir da correlação.

Tabela 3.4: Peso das Variáveis - Correlação

Variável	Número	Beta Correlação
Intercepto	0	4.2452
Casado	3	-0.34639
Nº Meses Adimp.	5	-0.65203
1-LTV	12	-1.8584
Valor Compra	14	-0.000025493
% Garantia	16	0.066086



Com os seguintes resultados:

Tabela 3.5: Resultados Correlação - Regressão Logística

Acurácia	78%
Erro	22%
Sensibilidade	71%
Precisão	82%
Especificidade	74%
KS	60%

## 3.6 SVM

Como o modelo de SVM sem kernel não possui solução, sabemos que os dados não são linearmente separáveis. Em nossos testes, os modelos de SVM não se saíram tão bem quanto os de regressão logística.

Para resolver o problema, utilizamos a função *fitcsvm* do MATLAB e testamos vários cenários. A função kernel polinomial não conseguiu resolver o problema e a função com kernel gaussiano se saiu melhor do que a com kernel linear. Contrariando o modelo anterior, as variáveis selecionadas por correlação mostraram mais ganho no caso do SVM.

Apresentaremos os resultados para os casos a seguir:

Tabela 3.6: Resultados Correlação - SVM com Kernel Gaussiano

Acurácia	68%
Erro	32%
Sensibilidade	83%
Precisão	33%
Especificidade	95%
KS	28%

Tabela 3.7: Resultados Correlação - SVM com Kernel Linear

Acurácia	32%
Erro	68%
Sensibilidade	18%
Precisão	16%
Especificidade	45%
KS	35%

É interessante observar que apesar da taxa de acerto do kernel linear ser muito menor do que a do kernel gaussiano, o KS é maior. Isto pode ser explicado pelo número pequeno de amostras de teste na carteira. Como existem poucos clientes maus, não podemos ter certeza de que o modelo está discriminando bem.

Tabela 3.8: Resultados Stepwise - SVM com Kernel Gaussiano

Acurácia	49%
Erro	51%
Sensibilidade	46%
Precisão	100%
Especificidade	9%
KS	13%

Tabela 3.9: Resultados Stepwise - SVM com Kernel Linear

Acurácia	35%
Erro	65%
Sensibilidade	30%
Precisão	38%
Especificidade	33%
KS	25%

## 3.7 VaR

Para calcular o VaR, obtivemos as informações históricas de pagamentos de um condomínio associado a empresa parceira. O condomínio começou a vender lotes em 2013, portanto temos informações desde junho de 2013 até março de 2018.

Com essas informações, calculamos a variação percentual entre os valores recebidos (R) e esperados (E) mês a mês.

$$\frac{R}{E} - 1$$

A partir daí, tivemos uma ideia como a carteira se comportou ao longo do tempo. O máximo de perda foi de 13% e o máximo de lucro foi 60%. Um fato isolado que ocorreu devido a quitação e adiantamentos no mesmo mês.

Para calcular o VaR histórico, utilizamos o Microsoft Excel, e obtivemos os seguintes resultados:

Com 95% de nível de confiança, a perda no próximo mês não será pior que: 8.41%

Em outras palavras, existe 5% de chance do retorno ser pior que -8.41%

O VaR paramétrico não apresentou muitas diferenças: Com 95% de nível de confiança, a perda não será pior que: 9.97%

Ou seja, existe 5% de chance do retorno ser pior que -9.97%

Nos casos de simulação, verificamos que a variação não passa de 20%. Ou seja, o modelo prevê que a perda e o ganho, não será maior que 20%.

Para o bootstrap, obtivemos que:

Com 95% de nível de confiança, a perda não será maior do que -5.605%

E como existem vários cenários, avaliamos o valor máximo e mínimo da perda:

VaR mínimo: -13.50%

VaR máximo: 6.93%

No último caso, que chamamos de simulação paramétrica, obtivemos os seguintes resultados:

Com 95% de nível de confiança, a perda não será maior do que 15.810%

Porém a variação ficou muito maior, devido ao intervalo que utilizamos para criar os cenários aleatórios. Obtivemos assim:

VaR mínimo: -71.14%

VaR máximo: 16.52%

Portanto, é necessário entender melhor o apetite de risco da empresa. Caso queiram ser conservadores, os modelos paramétricos são mais indicados.

# Capítulo 4

## Conclusões e trabalhos futuros

O objetivo deste trabalho foi ter uma visão inicial de como prever riscos para uma empresa que atua na área de securitização. Ao criar modelos comportamentais para os clientes, podemos ter uma ideia de quanto a carteira poderá perder se os clientes entrarem em default (passarem a ser inadimplentes).

De todos os modelos testados, o de regressão logística com variáveis selecionadas pelo método stepwise se destacou. O método stepwise, por verificar a variação de cada variável e também a correlação, apresenta resultados mais robustos, isto é, o modelo se torna mais estável. A regressão logística possui um método relativamente mais simples que o SVM e é muito mais utilizada em problemas relacionados a risco de carteira, portanto não foi uma surpresa que o método tenha se saído melhor. Como a empresa possui pouca informação sobre cada cliente, as variáveis dos modelos não apresentaram tanto ganho em relação ao SVM.

Já o método VaR se saiu muito bem em mesurar o risco mensal da carteira. Caso a empresa tenha um apetite de risco maior é aconselhado utilizar os métodos históricos. Caso contrário, os métodos paramétricos são mais conservadores.

Podemos aplicar o método do VaR para atualizar a curva de amortização mensalmente. Já os modelos de aprendizado de máquina podem ser utilizados para selecionar as carteiras elegíveis e estimar a perda total da carteira, bem como ter uma prévia de quais clientes poderão dar prejuízo e assim dispensar uma atenção maior a estes.

Reconhecemos que o modelo precisa ser testado em todas as carteiras e eventualmente será necessário criar um modelo geral, com uma carteira maior, para que os resultados sejam mais coerentes.

Também é necessário testar outras técnicas para categorizar dados e selecionar variáveis. Métodos como redes neurais também podem apresentar ganhos na separação entre clientes bons ou maus.

Futuramente, iremos acrescentar a simulação de Monte Carlo aos estudos e validar os modelos nos meses seguintes.

# Referências Bibliográficas

- [1] Ethos Group — Logos Securitizadora.  
<http://ethosgroup.com.br>
- [2] P. Morettin e W. Bussad *Estatística Básica* Editora Saraiva.
- [3] BM&FBOVESPA. Certificados de Recebíveis Imobiliários.  
[www.bmfbovespa.com.br/pt\\_br/.../certificados-de-recebiveis-imobiliarios.htm](http://www.bmfbovespa.com.br/pt_br/.../certificados-de-recebiveis-imobiliarios.htm)  
Consultado em 23/03/2018.
- [4] TESTE F - DUAS VARIÂNCIAS.  
[www.portalaction.com.br/manual-estatistica-basica/teste-f-duas-variancias](http://www.portalaction.com.br/manual-estatistica-basica/teste-f-duas-variancias)  
Consultado em 06/06/2018.
- [5] Stepwise regression.  
[www.mathworks.com/help/stats/stepwisefit.html](http://www.mathworks.com/help/stats/stepwisefit.html)  
Consultado em 06/06/2018.
- [6] Aprendizado de Máquina: Avaliação de Classificadores. **Revista Brasileira de Web:** Tecnologia.  
[www.revistabw.com.br/revistabw/avaliacao-de-classificadores/](http://www.revistabw.com.br/revistabw/avaliacao-de-classificadores/)  
Consultado em 03/06/2018.
- [7] Ebook: Machine Learning with MATLAB.  
[www.mathworks.com/campaigns/offers/machine-learning-with-matlab.html](http://www.mathworks.com/campaigns/offers/machine-learning-with-matlab.html)  
Consultado em 22/04/2018.
- [8] Revista Tecnologia de Crédito, Edição 93 - Modelos de Credit Scoring: uma Revisão Atualizada.
- [9] C. Diniz e F. Louzada. *Métodos Estatísticos para Análise de Dados de Crédito* 6th Brazilian Conference on Statistical Modelling in Insurance and Finance - Maresia/SP, 2013.

- [10] S. Chang, S. Kim e G. Kondo *Predicting Default Risk of Leading Club Loans* Stanford University, CS229: Machine Learning (Autumn 2015-2016).
- [11] E. Krulikowski. *Máquinas de Vetores de Suporte e Regressão Logística aplicados ao problema de análise de crédito* Universidade Federal do Paraná, 2016.
- [12] Y. Cheung e R. Powell *Anybody can do Value at Risk: A Nonparametric Teaching Study* Australasian Accounting, Business and Finance Journal.
- [13] D. Figueira. *Cálculo de VaR para uma carteira de ações: Sistema informatizado para ações negociadas na Bovespa* UFRJ, 2013.
- [14] Logistic Regression  
[en.wikipedia.org/wiki/Logistic\\_regression](http://en.wikipedia.org/wiki/Logistic_regression)
- [15] Distribuição Normal  
[www.portaaction.com.br/probabilidades/62-distribuicao-normal](http://www.portaaction.com.br/probabilidades/62-distribuicao-normal)
- [16] Support Vector Machine  
[en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine)
- [17] Teste KS  
[pt.wikipedia.org/wiki/Teste\\_Kolmogorov-Smirnov](http://pt.wikipedia.org/wiki/Teste_Kolmogorov-Smirnov)