

Você é o Melhor Classificador!

5 CONCEITOS BÁSICOS

PARA SAIR DO ZERO EM
MACHINE LEARNING



Autor do Livro Digital:

EDGAR MODESTO AMAZONAS

Livro digital criado por: Edgar Modesto Amazonas Filho – Todos os direitos reservados – www.amazonasfilho.com – Este produto eletrônico gratuito e não pode ser vendido. Proibida a reprodução ou alteração sem a autorização do seu criador!

SOBRE O AUTOR



Edgar Amazonas Filho é professor há 14 anos, engenheiro, programador, investidor e empreendedor. Participou em uma dezena de projetos para diversas concessionárias de Energia Elétrica do Brasil. Sempre usando o que há de mais avançado em técnicas computacionais: inteligência artificial e machine learning. Edgar acredita que muitos brasileiros podem aprender machine learning e aplicá-la, verdadeiramente, em problemas reais...e ajudando diversas empresas do Brasil.

Prefácio

Acredito que você tem ouvido cada vez mais sobre estas três áreas: inteligência artificial, machine learning e deep learning. Elas integram um tripé para o que vêm sendo citada como uma das profissões do futuro: CIÊNCIA DE DADOS. Essas três áreas envolvem transformar DADOS BRUTOS EM INFORMAÇÕES ÚTEIS PARA QUE UMA DECISÃO SEJA TOMADA DE FORMA AUTÔNOMA OU NÃO-AUTÔNOMA. Como as máquinas fazem isso vai além das pretensões deste livro introdutório.

Meu objetivo com esse livro é apenas um: mostrar que Machine Learning pode ser aprendida e aplicada por muitas pessoas.

À primeira vista você pode imaginar que Machine Learning é um assunto complexo e que apenas algumas pessoas “escolhidas” são capazes de dominá-la. Esse livro é para ajudar os “não-escolhidos”. Aqueles que querem aplicar em seu próprio trabalho acadêmico ou profissional.

Muitas vezes, a palavra “básico” é usada de forma errada... e até mesmo de forma pejorativa. Quem nunca ouviu a seguinte frase dita de forma desdenhosa: “Cara, isso é muito básico!”

O conhecimento básico, EM QUALQUER ÁREA, é simples, mas muito difícil de se executar perfeitamente. Somente quando você domina o básico, o mais complexo pode ser aprendido. O processo de aprender o básico é muito repetitivo, chato e frustrante. Eu, por exemplo, me apaixonei e aprendi dois hobbies: violão e taekwondo (uma arte marcial coreana). Quem já experimentou um desses hobbies sabe como eles são frustrantes no início e por isso a desistência nesta fase inicial é comum. E por quê? Porque todo mundo quer fazer aquele solo fantástico no violão...aquele chute complexo que o mestre faz maravilhosamente, mas ninguém quer treinar o básico...o chato.

Nesse livro, vou ensinar a você 5 conceitos básicos para começar a

aprender Machine Learning a partir do zero.

Eu separei esses conceitos com base em minha experiência de mais de 15 anos trabalhando nessa área. Escolhi esses 5 porque eles têm a principal característica do básico: simplicidade. Ao dominá-los você conseguirá aprender MUITO MAIS FÁCILMENTE as técnicas e modelos de Machine Learning. Mas você não pode esquecer de um importante fato: o básico é apenas a base para chegar no topo. Imagine algo como um arranha-céu, a base dele são suas fundações, e até que a cobertura seja construída e o prédio tenha o seu acabamento todo concluído, muita coisa tem que ser construída entre suas fundações e a sua cobertura.

Vou fazer uma última confissão aqui...eu nunca fui o aluno mais brilhante, mas poucos são tão determinados como eu. Lembro que no ensino médio o meu professor de física/matemática dizia bem assim: "...quem conseguir 4 D's (algo equivalente à média 5.0) nos quatro bimestres das minhas disciplinas, consegue passar na [universidade] federal". Quando eu concluí o ensino médio as minhas médias bimestrais foram 4 D's em física e 3 D's e um E (equivalente a menos que a média 5.0) em matemática...e mesmo com essas médias "medíocres" eu decidi que iria concorrer a uma vaga para o curso de engenharia elétrica na Universidade Federal do Pará, e...PASSEI!...em 115° de 120 vagas ofertadas...mas PASSEI. Isso já faz 22 anos, mas acredito que foi o primeiro momento em que descobri que TRABALHAR DURO, RESOLVE!

Todo MESTRE já foi um desastre.

(T. Harv Eker)

I. Introdução

Para definir os 5 conceitos básicos para você aprender Machine Learning eu perguntei a mim mesmo o que me faz entender mais facilmente os conceitos APARENTEMENTE complexos envolvidos em Machine Learning. Muito provavelmente você já estudou esses 5 conceitos, e se isso é verdade, se você VERDADEIRAMENTE os domina, então você já deu o primeiro passo para entender Machine Learning...apenas use esse livro para revisar esses conceitos e ainda conhecer alguns exemplos de suas aplicações. Mas se você ainda não os domina, então esse livro é para você...e nesse caso você precisa ler esse livro com muito MAIS ATENÇÃO.

Os **5 conceitos básicos** para você aprender mais facilmente Machine Learning são:

- 1. Conceito 1: Porcentagem**
- 2. Conceito 2: Funções**
- 3. Conceito 3: Geometria Analítica**
- 4. Conceito 4: Matriz**
- 5. Conceito 5: Somatórios e Produtórios**

Para cada um desses conceitos básicos eu vou colocar equações que realmente vão ser úteis para você entender Machine Learning, também vou incluir um exemplo envolvendo o uso do conceito aplicado em Machine Learning. Com essa estratégia você já saberá, antes mesmo de estudar Machine Learning, as principais equações básicas que serão úteis para o seu aprendizado sobre Machine Learning, e o mais importante você já saberá quando elas serão usadas. Dessa forma, mesmo que já você domine esses 5 conceitos, será importante fazer uma revisão rápida em cada um deles. Também, para cada um dos 5 conceitos você vai encontrar um exemplo e uma atividade para você fazer.

Além desses 5 conceitos básicos, há um **bônus** nesse livro:

Uma Introdução à Machine Learning.

Essa introdução aborda os tipos principais de aprendizagem usados em Machine Learning:

- **Supervisionado: regressão ou classificação**
- **Não-supervisionado: clustering**

2. Conceito I: Porcentagem

Dos 5 conceitos básicos, a **porcentagem** é o conhecimento mais simples. O valor da porcentagem P é obtido a partir de um total T para uma determinada parte p .

Para obter a porcentagem basta usar a equação (1) ou a equação (2). O resultado de (1) pode ser maior que 1 e de (2) pode ser maior que 100% para os casos em que o numerador p seja maior que o denominador T .

Equação Geral:

$$P = \frac{p}{T}$$

$$P = \frac{p}{T} \times 100 (\%)$$

Exemplo Aplicado em Machine Learning:

Machine Learning é uma ferramenta poderosíssima no processo de classificação. Para isso há diversos modelos classificadores, como exemplo: Regressão Logística¹, Máquina de Vetores de Suporte² (SVM), Árvores de Decisão³, e outros mais. Uma vez que esses modelos tenham sido treinados usando um conjunto de **dados de treino** ($\text{Entrada}_{treino}, \text{Saída}_{treino}$) eles podem ser usados para avaliar os seus desempenhos com base em conjuntos de **dados de teste** ($\text{Entrada}_{teste}, \text{Saída}_{teste}$). Agora imagine que você é um Cientista de Dados especialista em Machine Learning no Banco

¹ No inglês: *Logistic Regression*

² No inglês: *Support Vector Machine*

³ No inglês: *Decision Trees*

Itaú. Você precisa propor um modelo altamente eficiente para classificar se um novo cliente que está solicitando um financiamento de um automóvel vai dar calote ou não após obter o financiamento. O banco possui dados pessoais de 1485 clientes que já realizaram empréstimos em anos anteriores classificados como “caloteiros” e “não-caloteiros”. Por exemplo, você treinará três modelos: Regressão Logística, SVM, Árvores de Decisão treinados com e agora precisa avaliar o desempenho deles. Para isso você terá que avaliar a porcentagem de acertos de cada um dos modelos levando em consideração um conjunto de dados de teste.

Como o conceito básico de porcentagem é importante em Machine Learning?

i. Porcentagem de dados para treino e porcentagem de dados para teste

O total de clientes é 1485. A gente pode separar 80% de clientes para **dados de treino** e 20% de clientes para **dados de teste**. Então usando a formula da porcentagem:

Para os **dados de treino**, quantos clientes serão usados?

$$P_{treino} = \frac{p_{treino}}{T} \times 100$$

$$80 = \frac{p_{treino}}{1485} \times 100$$

$$p_{treino} = 1188 \text{ clientes}$$

Para os **dados de teste**, quantos clientes serão usados?

AGORA É COM VOCÊ MESMO!

ii. Porcentagem de saídas previstas x saída reais para o conjunto de dados de teste para três modelos de Machine Learning.

Para um dado conjunto de dados de entrada de teste⁴, os modelos de Machine Learning fornecem saídas previstas correspondentes. Por exemplo, considerando os resultados de 297 clientes na forma a seguir:

⁴ Esse tipo de teste é necessário antes que o modelo de Machine Learning seja definitivamente implantado em um sistema que transformará dados brutos em informações úteis para que seja feita alguma decisão.

Entrada	Saída Real	Saída Prevista	
Dados Cliente 1	Não-Caloteiro	Caloteiro	Erro Tipo I
Dados Cliente 2	Caloteiro	Caloteiro	
Dados Cliente 3	Caloteiro	Caloteiro	
Dados Cliente 4	Caloteiro	Não-Caloteiro	Erro Tipo II
:	:	:	
Dados Cliente 297	Não-Caloteiro	Não-Caloteiro	

Para se avaliar o desempenho dos modelos de Machine Learning, você precisa conhecer os seguintes termos:

P: número total de respostas positivas, ou seja, Saída Real [positiva] = ‘Caloteiro’

N: número total de respostas positivas, ou seja, Saída Real [negativa] = ‘Não-Caloteiro’

PV: número de respostas classificadas corretamente como positivas, ou seja, Saída Prevista [positiva] é igual a Saída Real [positiva]

NV: número de respostas classificadas corretamente como negativas, ou seja, Saída Prevista [negativa] é igual a Saída Real [negativa]

PF: número de respostas classificadas erroneamente como positivas, ou seja, Saída Prevista [positiva] é diferente da Saída Real [negativa]. Também conhecido como Erro Tipo I, ou Falso Alarme⁵, ou Falso Positivo.

NF: número de respostas classificadas erroneamente como negativas, ou seja, Saída Prevista [negativa] é diferente da Saída Real [positiva]. Também conhecido como Erro Tipo II, Engano⁶, ou Falso Negativo.

As métricas de desempenho usadas em Machine Learning [Classificação] são:

Taxa de Positivos de Verdadeiros:

$$TPV = \frac{PV}{P} = \frac{PV}{PV + NF}$$

⁵ A palavra ‘Alarme’ é usada porque em certos problemas – como esse de ‘calote’ em empréstimos – a Saída Positiva está relacionada a algo ruim ou perigoso, e que deve ser dada maior atenção por quem vai tomar a decisão final.

⁶ Em inglês, a palavra é ‘miss’ que há várias traduções possíveis, mas ‘engano’ está mais em linha com o que se deseja representar.

Taxa de Negativos de Verdadeiros:

$$TNV = \frac{NV}{N} = \frac{NV}{NV + PF}$$

Taxa de Positivos de Falsos:

$$TPF = \frac{PF}{PF + NV} = 1 - TNV$$

Taxa de Negativos de Falsos:

$$TNF = \frac{NF}{NF + PV} = 1 - TPV$$

Valores de Positivos Previstos:

$$VPP = \frac{PV}{PV + PF}$$

Valores de Negativos Previstos:

$$VNP = \frac{NV}{NV + NF}$$

Exatidão⁷:

$$E = \frac{PV + PN}{PV + PF + NV + NF}$$

Como calcular esses valores quando nos deparamos com os resultados dos modelos de Machine Learning?

⁷ Em inglês o termo vem da palavra ‘Accuracy’. Algumas traduções são feitas usando a palavra em português ‘Precisão’. Entretanto, ‘Exatidão’ é diferente de ‘Precisão’. Esta última envolve análise estatística.

Modelo de Regressão Logística:

		Saída Prevista	
		Caloteiro	Não-Caloteiro
Saída Real	Caloteiro	2 PV	79 NF
	Não-Caloteiro	17 PF	199 NV

As métricas de desempenho para os resultados do modelo de Regressão Logística:

Taxa de Positivos de Verdadeiros:

$$TPV = \frac{2}{2 + 79} = 0.025 = 2,469\%$$

Taxa de Negativos de Verdadeiros:

$$TNV = \frac{199}{199 + 17} = 0.921 = 92,13\%$$

Taxa de Positivos de Falsos:

$$TPF = \frac{17}{17 + 199} = 0.079 = 7,87\%$$

Taxa de Negativos de Falsos:

$$TNF = \frac{79}{79 + 2} = 0.975 = 97,531\%$$

Valores de Positivos Previstos:

$$VPP = \frac{2}{2 + 17} = 0.105 = 10,526\%$$

Valores de Negativos Previstos:

$$VNP = \frac{199}{199 + 79} = 0.716 = 71,583\%$$

Exatidão⁸:

$$E = \frac{2 + 199}{2 + 17 + 199 + 79} = 0.677 = 67,677\%$$

Modelo SVM:

		Saída Prevista		
		Caloteiro	Não-Caloteiro	
Saída Real	Caloteiro	78	PV	NF
	Não-Caloteiro	13	PF	NV
			203	

Calcule as métricas de desempenho para os resultados do modelo SVM.

AGORA É COM VOCÊ MESMO!

Modelo de Árvore de Decisão:

		Saída Prevista		
		Caloteiro	Não-Caloteiro	
Saída Real	Caloteiro	68	PV	NF
	Não-Caloteiro	55	PF	NV
			13	161

Calcule as métricas de desempenho para os resultados do modelo Árvore de Decisão.

AGORA É COM VOCÊ MESMO!

⁸ Em inglês o termo vem da palavra 'Accuracy'. Algumas traduções são feitas usando a palavra em português 'Precisão'. Entretanto, 'Exatidão' é diferente de 'Precisão'. Esta última envolve análise estatística.

3. Conceito 2: Funções

Eu me atrevo a dizer que o entendimento sobre funções é o conhecimento básico mais importante para COMEÇAR a entender Machine Learning, porque independentemente de qual técnica você for usar, de uma maneira ou de outra você estará em busca de uma função. Em resumo, Machine Learning também é a arte de tentar encontrar uma função que mapeie os valores de x em y . e essa a principal característica de uma função matemática.

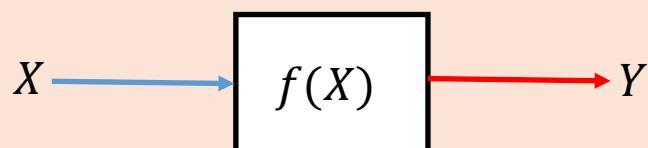
Equação Geral:

$$y = f(x)$$

Exemplo Aplicado em Machine Learning:

Em Machine Learning há duas formas chamadas de aprendizado: supervisionado e não-supervisionado. Esses dois tipos de aprendizados estão totalmente relacionados ao conhecimento básico de funções.

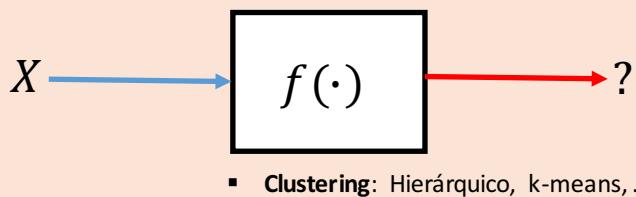
No aprendizado supervisionado nosso objetivo é ensinar a máquina de tal forma que ela consiga encontrar uma função f que relate a entrada X e a saída Y . São duas as aplicações do aprendizado supervisionado: REGRESSÃO e CLASSIFICAÇÃO. Esse aprendizado é chamado de supervisionado porque TEMOS que ensinar⁹ a máquina que X está relacionado a Y .



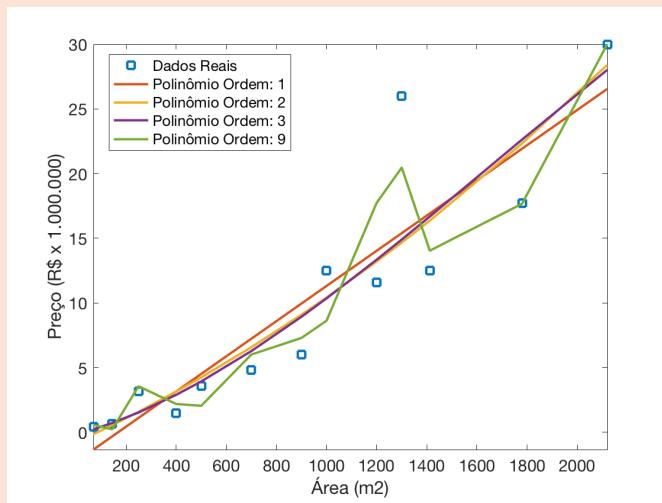
- **Regressão Paramétrica:**
Linear, Quadrática, Cúbica,...
- **Regressão Não-Paramétrica:**
Splines, Local, Kernel.
- **Classificação:** Regressão Logística, SVM, Árvores, ...

⁹ Esse ensino deve ser baseado em certezas, apoiado em dados confiáveis e cujas as respostas foram definidas por um especialista na área (administração, negócios, economia, engenharia, biologia, medicina, etc.).

No aprendizado não-supervisionado o objetivo não é ensinar a máquina, mas sim deixar que ela aprenda sozinha. O objetivo não é obter uma função como no aprendizado supervisionado, na verdade isso nem é possível, pois não se sabe a relação de X com Y . O que Machine Learning vai fazer é simplesmente agrupar todos os X que representam uma saída Y similar. Percebam que um próximo passo¹⁰ no aprendizado não-supervisionado é exatamente torná-lo em aprendizado supervisionado a partir da relação encontrada pelo processo de *clustering*.



Talvez nesse momento você esteja se perguntando: Afinal, o que pode ser X ou Y ? Essa é a chave de Ciência de Dados... são as variáveis. Então, o primeiro passo pra **implementar** Machine Learning é que você entenda bem o PROBLEMA que você deseja resolver e dessa forma consiga definir corretamente as suas VARIÁVEIS. A seguir um exemplo de **regressão** usando dados de imóveis do estado de São Paulo. Nesse exemplo, usei modelos de funções polinomiais, cujos parâmetros foram estimados usando técnicas estudadas em Machine Learning.



¹⁰ Se você ganhar habilidade aqui, você será extremamente valioso pra qualquer empresa e em qualquer setor.

Usando as funções estimadas usando Machine Learning podemos estimar o **preço** para a **área** de 2000 m².

função	equação	área (m ²)	preço (R\$xmilhão)
1º-grau	preço = 0.0136 · área – 2.2863	2000	24,914
2º-grau	preço = $2,29 \cdot 10^{-6} \cdot \text{área}^2 + 0.0089 \cdot \text{área} - 0.7809$	2000	26,179
3º-grau	preço $= -1.33 \cdot 10^{-9} \cdot \text{área}^3 + 6,63 \cdot 10^{-6} \cdot \text{área}^2 + 0.0052 \cdot \text{área} - 0.1686$	2000	26,111
9º-grau	preço $= -1.66 \cdot 10^{-24} \cdot \text{área}^9 + 1.4 \cdot 10^{-20} \cdot \text{área}^8 - 4,92 \cdot 10^{-17} \cdot \text{área}^7 + 9,34 \cdot 10^{-14} \cdot \text{área}^6 - 1,04 \cdot 10^{-10} \cdot \text{área}^5 + 7,02 \cdot 10^{-8} \cdot \text{área}^4 - 2,77 \cdot 10^{-5} \cdot \text{área}^3 + 0,006 \cdot \text{área}^2 - 0.6102 \cdot \text{área} + 21.85$	2000	41.30

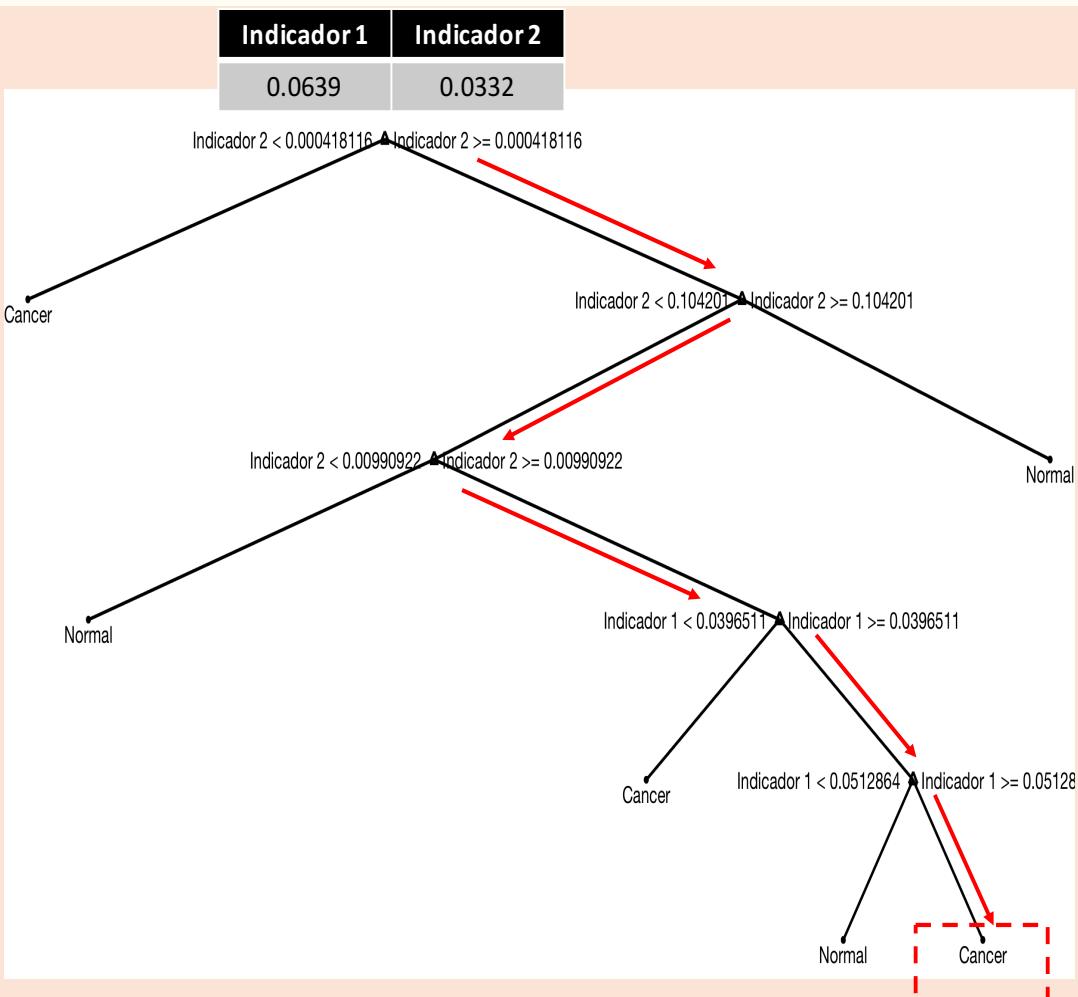
Calcule o preço estimado para um imóvel de 3250 m²

função	equação	área (m ²)	preço (R\$xmilhão)
1º-grau	preço = 0.0136 · área – 2.2863	3250	?
2º-grau	preço = $2,29 \cdot 10^{-6} \cdot \text{área}^2 + 0.0089 \cdot \text{área} - 0.7809$	3250	?
3º-grau	preço $= -1.33 \cdot 10^{-9} \cdot \text{área}^3 + 6,63 \cdot 10^{-6} \cdot \text{área}^2 + 0.0052 \cdot \text{área} - 0.1686$	3250	?
9º-grau	preço $= -1.66 \cdot 10^{-24} \cdot \text{área}^9 + 1.4 \cdot 10^{-20} \cdot \text{área}^8 - 4,92 \cdot 10^{-17} \cdot \text{área}^7 + 9,34 \cdot 10^{-14} \cdot \text{área}^6 - 1,04 \cdot 10^{-10} \cdot \text{área}^5 + 7,02 \cdot 10^{-8} \cdot \text{área}^4 - 2,77 \cdot 10^{-5} \cdot \text{área}^3 + 0,006 \cdot \text{área}^2 - 0.6102 \cdot \text{área} + 21.85$	3250	?

AGORA É COM VOCÊ MESMO!

A seguir temos um exemplo de **classificação** usando modelos de Machine Learning. O problema envolve classificar uma pessoa baseada em indicadores, cujo resultado será: câncer ou normal. Para isso vamos usar um modelo chamado Árvore de Decisão com 2 indicadores¹¹ para exemplificar.

¹¹ O problema original trabalha com 4000 indicadores, mas para fins de simplificação usei apenas 2 indicadores.



Indicador 1	Indicador 2	Resultado
0.0639	0.0332	Cancer
0.0254	0.0511	?
0.0255	0.0361	?
0.0128	0.0297	?
0.0198	-0.0106	?
0.0390	0.0394	?
0.0232	0.0538	?
0.0270	0.0101	?
0.0394	0.0389	?
0.0134	0.0165	?

Determine o resultado: Câncer ou Normal para os outros valores dos indicadores 1 e 2.

AGORA É COM VOCÊ MESMO!

4. Conceito 3: Geometria Analítica

O conhecimento básico de geometria analítica é especialmente importante em Machine Learning porque para ilustrar (comunicar) as estratégias adotadas em um problema, sempre vamos precisar a plotagem de gráficos, por exemplo, muitos cientistas de dados vão usar *scatter plots* – seja um problema de regressão, de classificação ou clustering – mostrando a distribuição de dados de entrada e relacionando-os as saídas. Esses dados, geralmente, vão ser plotados usando gráficos em duas dimensões e com os conjuntos de pares ordenados (X,Y) dispostos no plano cartesiano.

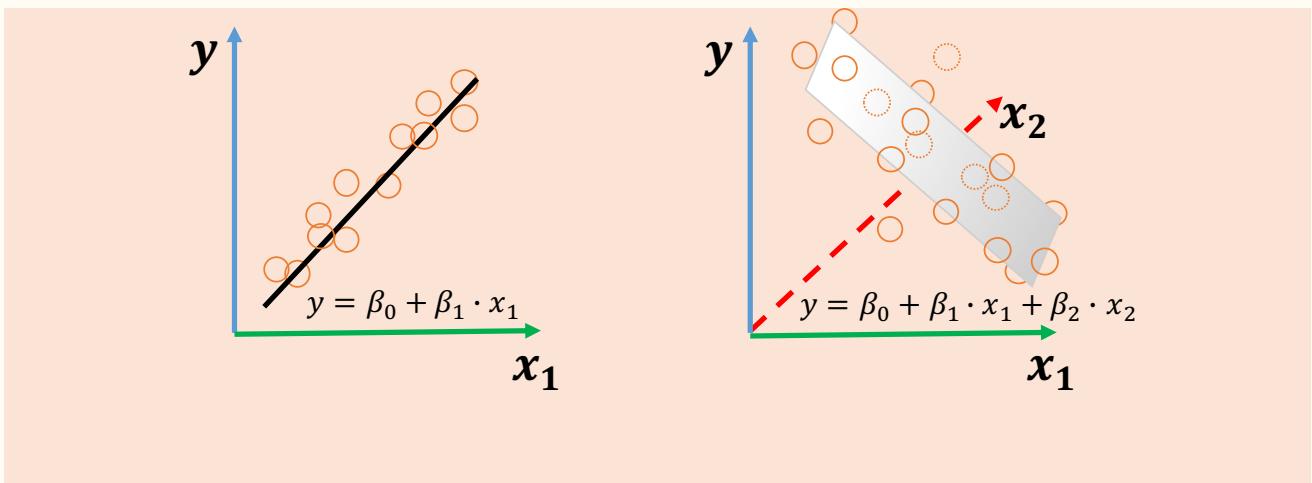
Equação Geral:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \cdots + \beta_p \cdot x_p$$

Exemplo Aplicado em Machine Learning:

Uma aplicação direta de geometria analítica é a Regressão Linear Simples ou Múltipla¹². Se tivermos trabalhando com uma variável de entrada, e uma de saída então precisamos ajustar uma equação de reta aos dados dispostos em 2D (duas dimensões). Se tivermos trabalhando com duas variáveis de entrada, e uma de saída então precisamos ajustar uma equação de plano aos dados em 3D (três dimensões)...além de duas variáveis não conseguiremos visualizar, e os dados serão ajustados ao que se chama de *hiperplano*.

¹² Muitas vezes você pode encontrar esse tópico, simplesmente, em estatística, ou outra área conhecida como Statistical Learning. Pra não perdermos o foco, sempre vou evitar esse tipo de discussão, e sempre adotarei o termo Machine Learning



Uma outra equação bem conhecida em Geometria Analítica é aquela que mede a distância entre dois pontos – a *distância euclidiana*. Apesar de simples, possui uma importantíssima aplicação em Machine Learning.

Equação Geral:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

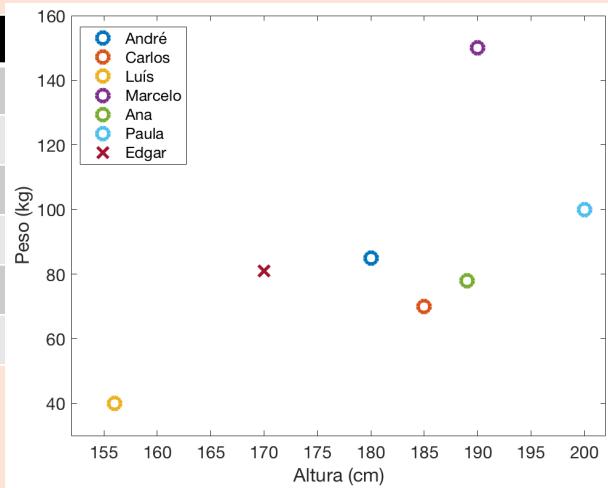
Exemplo Aplicado em Machine Learning:

Relembrando, no aprendizado não-supervisionado de Machine Learning nós fornecemos à máquina os dados sem sabermos uma relação clara que relacionam os dados de entrada e os dados de saída. Apesar de sua simplicidade, eu acredito que esse tipo de aprendizado é o mais desafiador de se aprender para os iniciantes em Machine Learning – afinal, como uma máquina pode aprender “sozinha”? ...apesar de parecer algo “surreal”, a verdade é que não há muito segredo nesse método conhecido como *clustering*. O objetivo da máquina é apenas um...juntar os SEMELHANTES ...e para isso, precisamos de alguma medida que diga o quanto um dado é semelhante a outros. Uma medida possível¹³ é exatamente a distância euclidiana.

¹³ Há outras mais sofisticadas, mas acredito que se você entender o fundamento por trás da medida estuda aqui, as outras se tornarão bem mais fáceis.

Vamos agora visualizar isso na prática e aplicada a princípios usados em Machine Learning. Calculando a distância entre mim (Edgar, Altura: 170 cm, Peso: 81 kg) e outras pessoas.

Nome	Altura (cm)	Peso (kg)	Distância
André	180	85	10,77
Carlos	185	70	?
Luís	156	40	?
Marcelo	190	150	?
Ana	189	78	?
Paula	200	100	?



Qual a distância entre mim e Carlos, Luís, Marcelo, Ana e Paula?

AGORA É COM VOCÊ MESMO!

5. Conceito 4: Matriz

Lembro quando estudei matrizes pela primeira vez...estava no meu 2º ano do ensino médio e pensava comigo mesmo – no “alto” da minha sabedoria adolescente: “Pra quê isso!...não serve pra nada!”... Só fui entender o porquê anos depois no meu curso de engenharia...lá “tudo” são matrizes. E alguns anos mais tarde, ao estudar e aplicar Machine Learning, novamente, foi reforçada a importância desse conhecimento básico fundamental pra muitas coisas em Machine Learning.

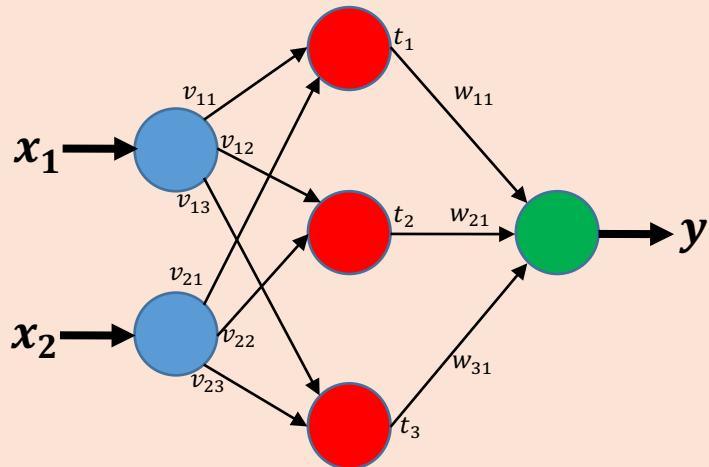
Uma matriz é representada da seguinte forma:

Equação Geral:

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

Exemplo Aplicado em Machine Learning:

Redes Neurais Artificiais (RNA) é uma técnica de Machine Learning¹⁴ usada tanto em Regressão como em Classificação. RNA usa o aprendizado supervisionado, ou seja, temos que fornecer um conjunto de dados relacionando entrada e saída, algo do tipo $X \rightarrow Y$. Abaixo temos uma RNA¹⁵ com 2 neurônios na camada de entrada, 3 neurônios na camada escondida (intermediária) e 1 neurônio na camada de saída. Apesar dos termos neurônio, rede neural, camadas etc. soarem como algo de ficção científica, na verdade após o processo de treinamento, a RNA será resumida, em grande parte, a multiplicações matriciais entre as camadas de neurônios: entrada, escondidas e saída.



As saídas dos neurônios da camada escondida podem ser calculadas usando a operação matricial de multiplicação:

$$\mathbf{T} = \mathbf{V} \cdot \mathbf{X}$$

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \\ v_{13} & v_{23} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Então, as saídas são¹⁶:

¹⁴ Alguns podem alegar que Redes Neurais Artificiais é uma técnica de Inteligência Artificial. Eu, particularmente, sempre vou evitar esse tipo de discussão, pois acredito que o Cientista de Dados deve focar mais em dominar as ferramentas e resolver problemas.

¹⁵ Em uma RNA usada em Machine Learning há alguns elementos a mais que não foram representados aqui apenas por simplificação.

¹⁶ Os índices usados nas matrizes se referem aos neurônios e não as posições dos elementos das matrizes propriamente ditos.

$$t_1 = v_{11} \cdot x_1 + v_{21} \cdot x_2$$

$$t_2 = v_{12} \cdot x_1 + v_{22} \cdot x_2$$

$$t_3 = v_{13} \cdot x_1 + v_{23} \cdot x_2$$

Escreva as equações matriciais para a saída y considerando a matriz de peso W

AGORA É COM VOCÊ MESMO!

6. Conceito 5: Somatórios e Produtórios

Chegamos ao último conceito básico para você começar a entender Machine Learning. Somatório e Produtório são conceitos simples advindos de cálculo numérico. Já o ensinei em outros cursos que ministrei e percebi que o principal problema que os iniciantes enfrentam é quando os elementos que vão estar no somatório ou no produtório se tornam algo além de uma simples soma ou produto de x_i elementos.

Equação Geral:

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 \dots + x_n$$

$$\prod_{i=1}^n x_i = x_1 \times x_2 \times x_3 \dots \times x_n$$

Exemplo Aplicado em Machine Learning:

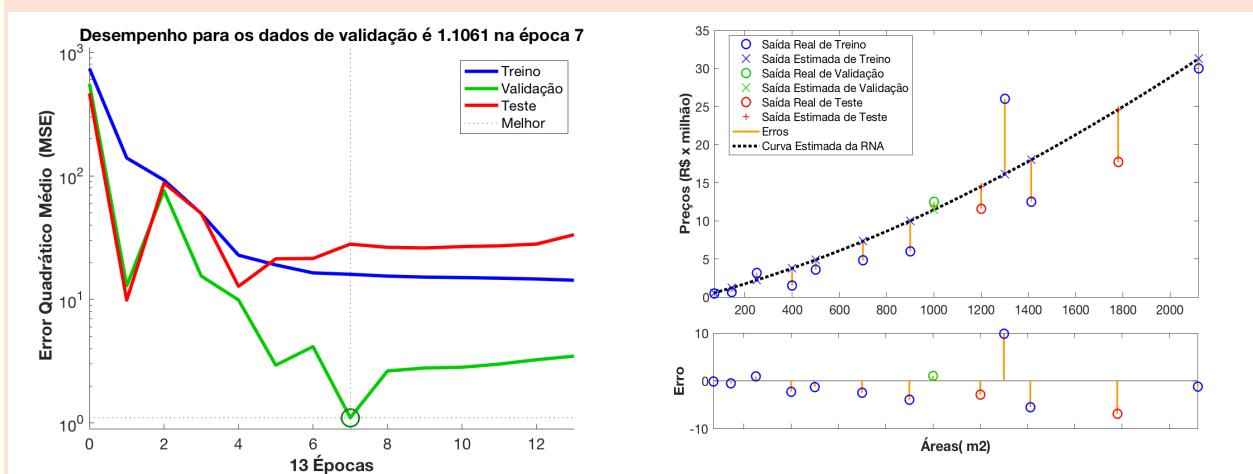
Vou aproveitar o conhecimento de somatórios para reforçar¹⁷ pra você o conceito MAIS IMPORTANTE pra você propor o melhor modelo de Machine Learning usando

¹⁷ Esse conceito foi abordado no primeiro exemplo

aprendizado supervisionado. Domine esse conceito e você sempre ficará confiante quando exibir os seus resultados a outras pessoas: ERRO DE TREINAMENTO e ERRO DE TESTE.

Em Machine Learning, quando usamos aprendizado supervisionado – regressão ou classificação – se quisermos propor um modelo confiável, vai ser obrigatório separarmos o nosso conjunto de dados em parte para o treinamento e uma outra parte para teste.

Quando você testar os seus modelos, então você poderá gerar um conjunto de gráficos que apoiarão o modelo “vencedor”. Como exemplo, os resultados abaixo eu obtive ao treinar uma Rede Neural Artificial (1 Neurônio na camada de entrada, 2 neurônios na camada escondida e 1 neurônio na camada de saída) pra resolver o problema do exemplo 2 (Precificação de Casas em Condomínios em São Paulo). Lembrando: o conjunto de dados (área e preço) consiste de 13 imóveis reais obtidos de um site de venda de imóveis. Esses dados foram divididos da seguinte forma: 10 móveis para treino, 1 imóvel para validação¹⁸ e 2 para testes.



O exemplo Preço X Área de casas em condomínios envolve um problema de regressão, ou seja, eu desejo prever a saída y com base na entrada x . O modelo RNA

¹⁸ O conjunto de dados de validação funcionam de forma semelhante ao conjunto de testes. Eles não fazem parte do treinamento, mas sim são testados durante o treinamento. Ele indicará a melhor solução para o modelo.

deve propor uma função (linha preta pontilhada acima) que faça isso. Para propormos o melhor modelo RNA, nós devemos considerar os erros MSE¹⁹: treino e teste (incluindo validação). Os erros são obtidos da seguinte forma:

Em regressão:

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i)]^2$$

x_1 Área	y Preço	\hat{y} Preço
70	0,430	0,405
145	0,655	0,700
250	3,180	2,785
400	1,500	3,100
500	3,600	4,450
700	4,800	6,150
900	6,000	7.670
1000	12,500	13,280
1200	11,600	14,000
1300	26,000	15,950
1412	12,500	18,760
1780	17.700	20,800
2120	30,000	31,120

Qual o erro MSE para estes resultados?

AGORA É COM VOCÊ MESMO!

Em classificação:

$$ME = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i)$$

¹⁹ Do inglês, MSE: Mean Squared Error, e em português, Erro Quadrático Médio. Em regressão, o valor previsto pode ser numericamente maior que o valor real, então por isso a necessidade de elevar ao quadrado a diferença entre valor previsto e valor estimado. Em classificação, não existe nenhuma operação de subtração, então não há necessidade de elevar ao quadrado a diferença entre a classe prevista e a classe real.

x_1 Indicador 1	x_2 Indicador 2	y Resultado	\hat{y} Resultado
0.0639	0.0332	Cancer	Cancer
0.0254	0.0511	?	Normal
0.0255	0.0361	?	Normal
0.0128	0.0297	?	Cancer
0.0198	-0.0106	?	Cancer
0.0390	0.0394	?	Cancer
0.0232	0.0538	?	Normal
0.0270	0.0101	?	Normal
0.0394	0.0389	?	Normal
0.0134	0.0165	?	Cancer

Qual o erro ME para estes resultados - use as respostas encontradas no exemplo 2?

AGORA É COM VOCÊ MESMO!

7. BÔNUS: Introdução a Machine Learning

Agora que você viu (ou reviu) 5 conceitos básicos para começar a aprender Machine Learning, então agora vou fazer um pequeno resumo dessa fascinante área que promete revolucionar todas as empresas de qualquer lugar do mundo.

Machine Learning na verdade está inserida dentro de um contexto bem maior chamado de Ciência de Dados. Pra gente entender melhor a Ciência de Dados, podemos visualizá-la como um processo que se inicia da aquisição de dados (brutos), que após um possível pré-processamento, vão ser enviados a um sistema (inteligente) que será capaz de gerar uma informação útil, e que por fim será usada para uma tomada de decisão – autônoma ou não-autônoma. Na decisão autônoma o próprio hardware executa alguma ação onde esse processo está se desenvolvendo, um exemplo dessa situação são os carros autônomos, por exemplo, os sensores captam as imagens que os cercam – pessoas, estradas, placas de trânsito, outros automóveis etc. – nesse caso o carro agirá acelerando, freando, parando, manobrando de forma totalmente autônoma. A decisão não-autônoma é a mais popularmente aplicada porque nela a ação é executada por uma pessoa, ou seja, a informação é apenas um elemento de auxílio na tomada de decisão, um exemplo dessa situação poderia ser um sistema de avaliação de crédito bancário. Nessa situação o bancário é que tomará a decisão final de conceder ou não o empréstimo – apesar de que na maioria das vezes o bancário irá simplesmente seguir a decisão oferecida pelo sistema (inteligente).



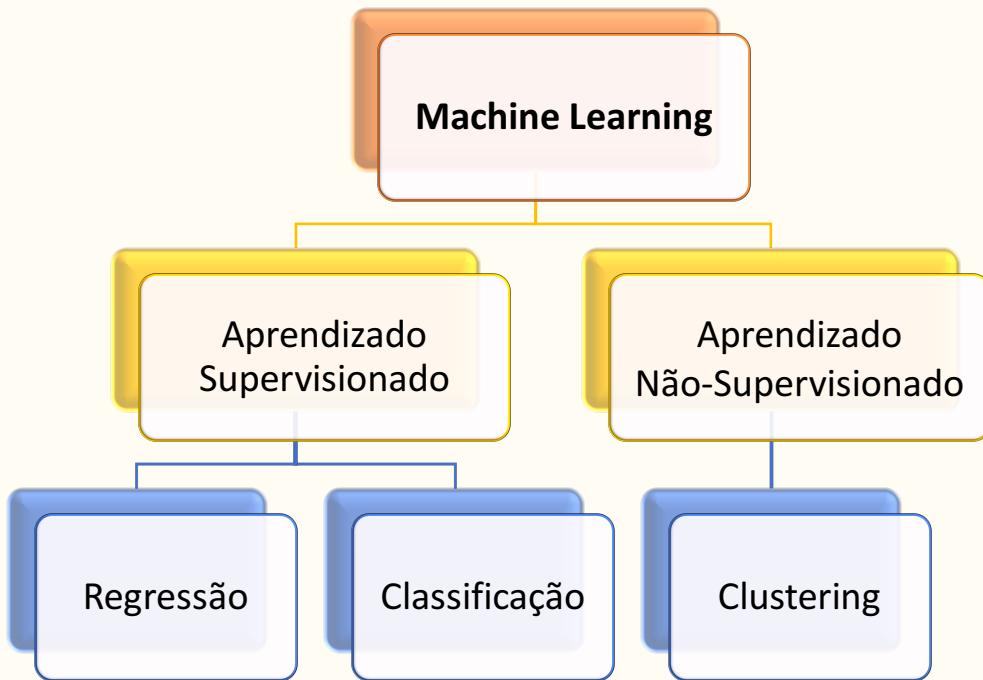
Machine Learning está situada exatamente no sistema (inteligente). Ele não necessariamente está sozinho ali... e na verdade em um sistema (inteligente) podem haver outras técnicas trabalhando em conjunto com Machine Learning, como: Inteligência Artificial e Deep Learning.

Relembrando a você... este livro não é para aprender Machine Learning, mas vou fornecer uma pequena introdução.

Traduzindo para o português Machine Learning significa Aprendizado de Máquina²⁰. As técnicas estudadas em Machine Learning podem ser divididas em dois tipos de aprendizados²¹: *supervisionado* e *não-supervisionado*. No aprendizado supervisionado, você (ou um especialista no problema que está sendo estudado) “ensina” a máquina a seguinte relação: *olha máquina, quando for apresentado como entrada ISSO, então a saída tem que ser ESSA*. Uma vez que a máquina “aprende”, então ela será capaz de responder sozinha quando for apresentada a ela uma nova entrada (não necessariamente idêntica àquelas usadas na fase de aprendizado). Já o aprendizado não-supervisionado a própria máquina é responsável por “aprender”. É como se você dissesse para a máquina assim: *as entradas são isso, então encontre conjuntos que sejam similares*. Esses conjuntos são chamados de *clusters*.

²⁰ Na minha opinião, usar a forma traduzida de Machine Learning soa tão estranha quanto chamar de ADN para DNA. Então, como dica, dê preferência para a forma em inglês ;-)

²¹ Ainda existem dois outros tipos também estudados em Machine Learning: semi-supervisionado e reforçado.



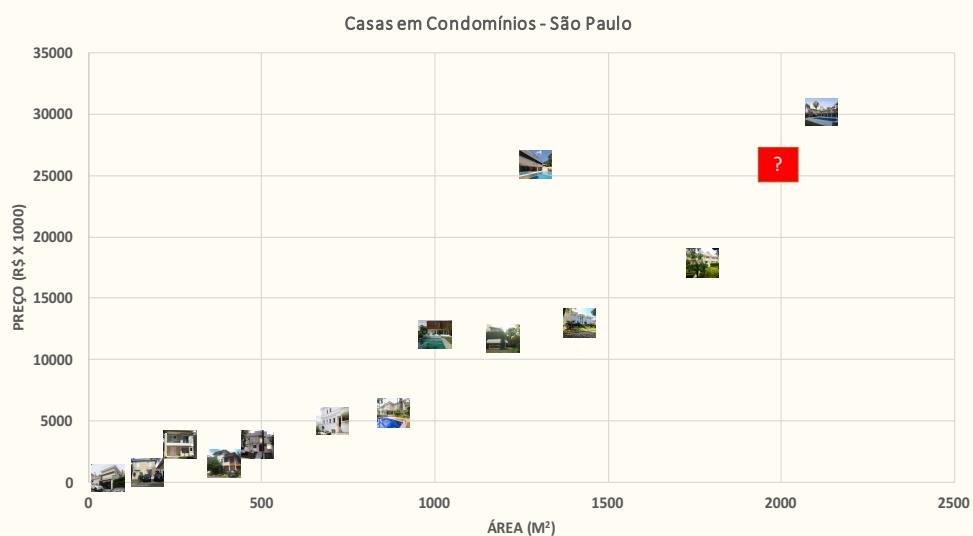
Eu vou usar exemplos para ilustrar os dois tipos de aprendizados utilizados em Machine Learning. São 3 exemplos: 2 envolvendo aprendizado supervisionado (um de regressão e 1 de classificação) e 1 envolvendo aprendizado não-supervisionado.

Exemplo 1: Aprendizado Supervisionado (Regressão)

Esse é um problema clássico em Machine Learning – Previsão de Preços de Imóveis. Nesse exemplo eu consultei um famoso site de anúncios imóveis, então escolhi 14 deles no estado de São Paulo e seus respectivos preços e áreas. Machine Learning é usada nesse tipo de problema com o objetivo de estimar uma função²² que relate área (eixo-x) e preço (eixo-y) dos imóveis. Para isso, deve-se fornecer um conjunto de pares (área, preço) para se fazer o processo de treinamento – por essa razão é que se chama aprendizado supervisionado. *Uma vez que a função seja estimada, então será possível usá-la para prever o preço de imóveis conforme uma área fornecida pelo usuário, por exemplo, a máquina será capaz de informar o preço de um imóvel cuja área é 2000 m².*

²² No processo real, Machine Learning é usada para estimar os coeficientes da função, e não a função em si.

Área (m ²)	Preço (R\$ x 1000)
70	430
145	655
250	3.180
400	1.500
500	3.600
700	4.800
900	6.000
1000	12.500
1200	11.600
1300	26.000
1412	12.500
1780	17.700
2120	30.000



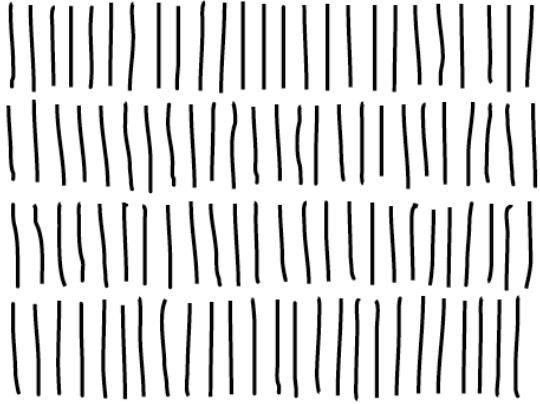
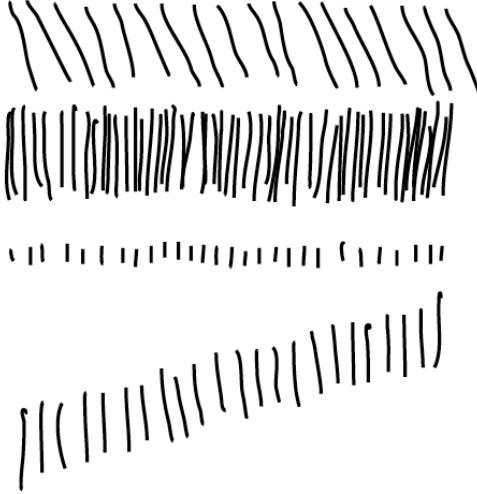
Exemplo 2: Aprendizado Supervisionado (Classificação)

Nesse exemplo tive a ideia de mostrar para você o potencial de Machine Learning alcançar outras áreas do conhecimento. Muitas pessoas como você já fizeram (ou fará) o “temeroso” teste psicotécnico do DETRAN para se habilitar no trânsito. Possivelmente, você se deparou (ou vai se deparar) com a questão dos “pauzinhos”. O objetivo dessa questão reconhecer padrões de comportamentos humanos²³, e reconhecer padrões Machine Learning também é fantástica. Sempre que houver um problema de reconhecimento de padrões com o objetivo de classificação, então haverá SEMPRE a necessidade de um especialista. Na questão dos “pauzinhos” o especialista é um psicólogo. Se fossemos criar uma solução²⁴ baseada em Machine Learning para corrigir, automaticamente, as questões dos candidatos. Para isso, o especialista terá que fornecer conjuntos de padrões normais e anormais, bem como as características (do candidato: idade, escolaridade, etc., dos “pauzinhos”:

²³ Esse exemplo é meramente ilustrativo e não serve como guia para desempenhar melhor em testes psicotécnicos. Também, não estamos discutindo o que é ser “Normal” ou “Anormal”. Apenas especialistas da área podem classificar mais corretamente os resultados obtidos nesse tipo de teste.

²⁴ Uma solução nesse caso poderia ser um app pra tablet, em que o usuário pudesse desenhar os “pauzinhos” na própria tela do tablet.

quantidade, distância, agrupamento, inclinação, tamanho, direção etc.) que foram usadas para a classificação. Esse problema é similar ao problema do Exemplo 1, pois novamente estamos fornecendo para a máquina um conjunto de pares (características, padrões) para se fazer o processo de treinamento – por essa razão é que se chama aprendizado supervisionado. *Também similarmente ao Exemplo 1, teremos uma função (não rigorosamente matemática) que será capaz de mapear características em padrões.* Essa função também é conhecida como **classificador**. Aparentemente, o classificador substituirá o especialista, mas se formos observar mais atentamente, o que se construiu foi um elemento de apoio à decisão do especialista.

Padrões Normais	Padrões Anormais
	

Exemplo 3: Aprendizado Não-Supervisionado (Clustering)

Nos dois exemplos anteriores mostrei que o aprendizado supervisionado sempre envolve uma fase chamada de treinamento. Nessa fase um especialista sempre deverá fornecer um conjunto de pares que se relacionam. Esses pares podem ser entendidos como: (X,Y), (Entrada, Saída), (Características, Classes), (Preditor, Resposta). No aprendizado não-supervisionado isso muda. Agora, apenas um lado do par é fornecido: os X's, as entradas, as características. Podemos ilustrar essa situação através de curvas de consumo de energia elétrica de diversos consumidores residenciais. Nesse experimento foram capturadas 96 medidas de consumo de energia elétrica ao longo de 24 horas, ou seja, a cada 15 minutos o consumo é armazenado²⁵ em uma base de dados. Esse procedimento foi repetido para 311 consumidores residenciais. Uma vez que temos a base de dados com essas 311 curvas, então podemos usar Machine Learning aplicando técnicas (*clustering*²⁶) baseadas em aprendizado não-supervisionado. Uma vez que esses dados sejam fornecidos, as técnicas de *clustering* são capazes de separar as curvas semelhantes sem que qualquer especialista intervenha no processo, ou seja, a máquina aprende sozinha...de forma não-supervisionada. Olhando a figura 1 você pode ver esses *clusters*²⁷ que se formaram. As curvas de consumo de energia elétrica representadas por cada um desses *clusters* podem ser vistas na figura 2.

²⁵ Esse experimento pode ser relativamente fácil de ser reproduzido através de medidores inteligentes de energia elétrica.

²⁶ Em português, *clustering* significam aglomeração.

²⁷ Clusters são os aglomerados de dados que são semelhantes e que foram identificados usando *Machine Learning* com aprendizado não-supervisionado. Para visualização dos *clusters* eu estou usando a Análise de Componentes Principais. Essa é uma técnica largamente usada em *Machine Learning*, mas que está fora do escopo desse livro.

Cluster	Curvas Agregadas
1	87
2	35
3	9
4	7
5	138
6	3
7	2
8	27
9	3

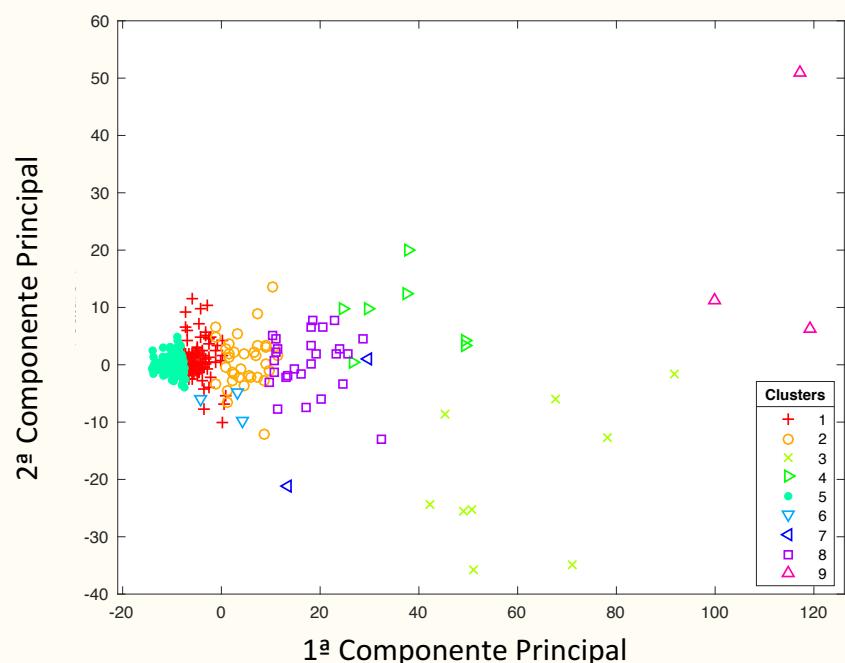
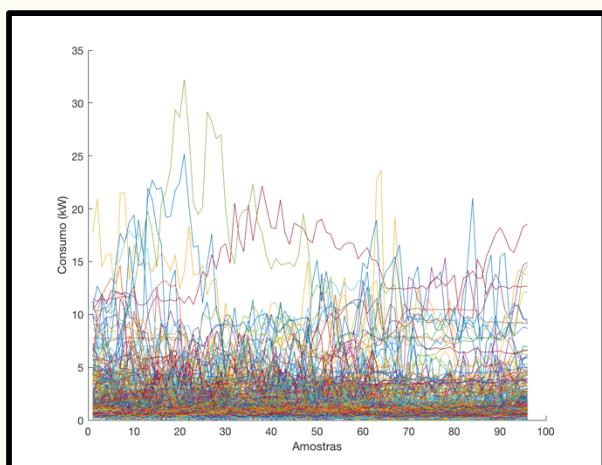
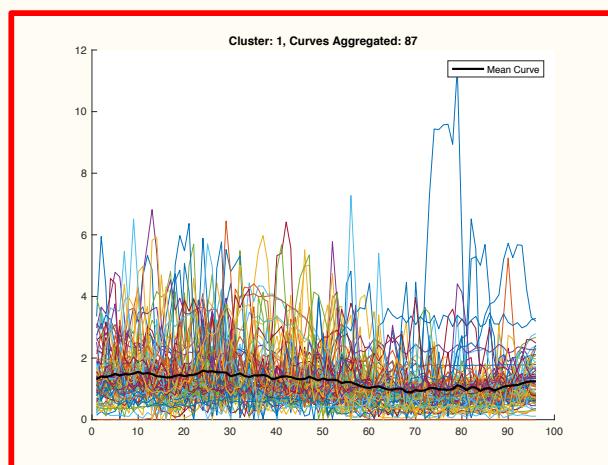


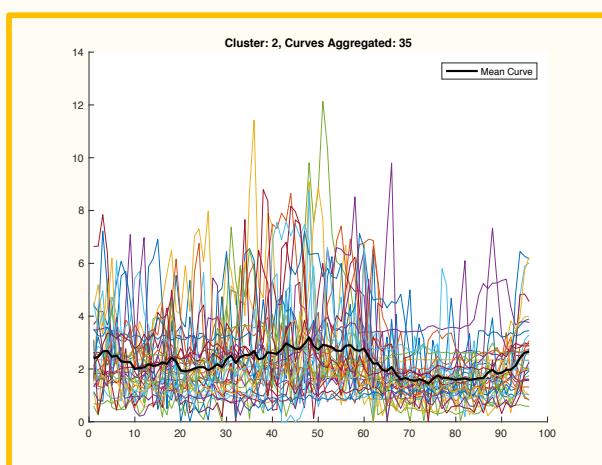
Figura 1



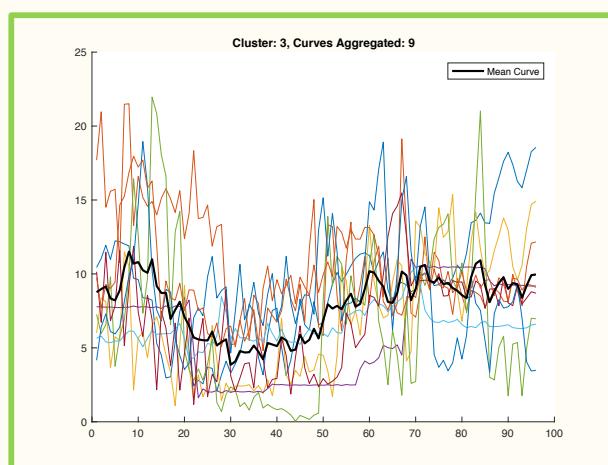
(a)



(b)



(c)



(d)

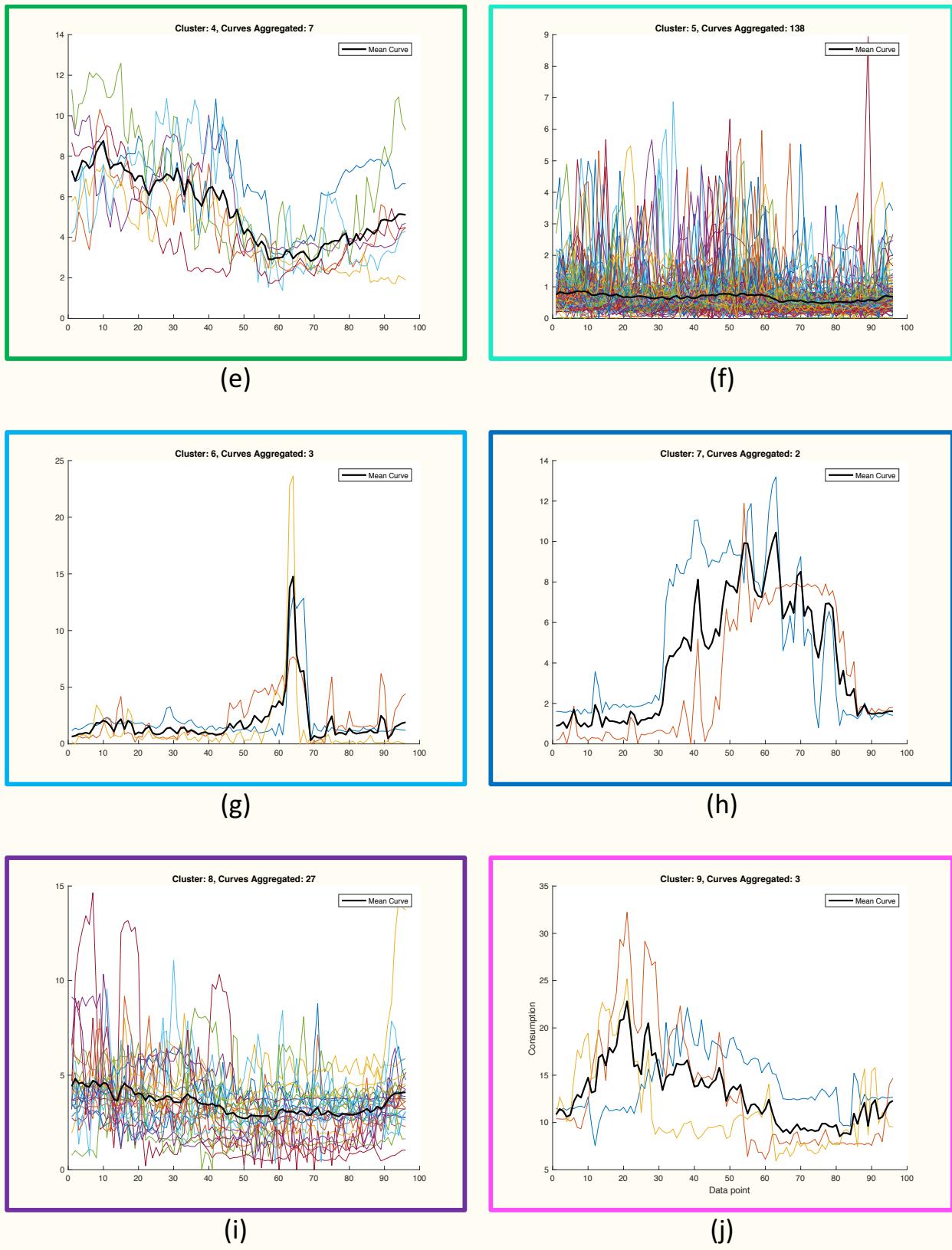


Figura 2

Eu sei que Machine Learning pode assustar à primeira vista, mas com esse livro eu decidi resumir os principais conceitos básicos para que não haja desânimo nesse primeiro encontro com essas técnicas novas. Com esse livro estou criando um pequeno atalho que eu acredito que vai ajudar você a evitar todas as dificuldades que eu mesmo tive ao longo do meu aprendizado em Machine Learning. Os 5 conceitos básicos devem ser completamente dominados antes de você seguir para os passos seguintes. E quais esses próximos passos? Poderia adiantar alguns: Álgebra Linear, Probabilidade, Estatística, Teste de Hipóteses e Otimização. Possivelmente, esses próximos passos você já estudou (ou estudará) em cursos de graduação e/ou pós-graduação. Mas o importante é que nesses próximos passos você se dedique apenas a pontos essenciais (caso contrário você vai perder tempo) para entender futuramente os modelos usados em Machine Learning.

Eu, verdadeiramente, acredito que Machine Learning possa ser aprendida por muitas pessoas. Esse conhecimento não deve ser limitado apenas a um nicho de pessoas. Espero que esses primeiros 5 passos façam com que você acredite que possa aprender Machine Learning. Lembro quando tinha 9 anos – e hoje tenho 39 anos – ouvia muito a frase: “O Brasil é o país do futuro”... e ainda ouço isso nos dias de hoje. Machine Learning é o futuro e se nós mesmos não formos para a AÇÃO, novamente, daqui a 30 anos você ainda vai ouvir a frase: “O Brasil é o país do futuro”.



Se dados é o novo petróleo, então Machine Learning é a nova refinaria.

Edgar Amazonas Filho