

Machine Learning Aplicada à Agropecuária de Precisão

Dia 01 - Machine Learning



Giancarlo D. Salton, PhD

Applied Intelligence Research Centre & ADAPT Centre
School of Computing

Chapecó, 08 de maio de 2018

Inteligência Artificial

Machine Learning

Métodos baseados em informação

Métodos baseados em similaridade

Métodos baseados em probabilidade

Métodos baseados em erro

Resumo

Inteligência Artificial

- ▶ Na Ciência da Computação, é o estudo de “agentes inteligentes” em um determinado domínio para atingir algum objetivo

- ▶ Em geral, os problemas podem ser resumidos em 2 tipos:
 - ▶ classificação
 - ▶ regressão

- ▶ A forma mais simples é através de um conjunto de regras que formam um modelo do domínio
- ▶ Modelos são aplicáveis apenas a um problema dentro do domínio

► Regras devem ser definidas por alguém!

- prolog
- banco de dados
- código fonte
- ...

ID	Profissão	Idade	Salário-Emprestimo	Proporção
				Classe
1	indústria	34	2.96	diferenciada
2	autônomo	41	4.64	padrão
3	autônomo	36	3.22	padrão
4	autônomo	41	3.11	padrão
5	indústria	48	3.80	padrão
6	indústria	61	2.52	diferenciada
7	autônomo	37	1.50	diferenciada
8	autônomo	40	1.93	diferenciada
9	indústria	33	5.25	padrão
10	indústria	32	4.15	padrão

```
if Proporção Salário-Empréstimo > 3 then
    Classe= 'padrão'
else
    Classe= 'diferenciada'
end if
```

ID	Valor	Renda Anual	Razão		Profissão	Propriedade	Classe
			Renda Empr.	Idade			
1	245,100	66,400	3.69	44	indústria	fazenda	diferenciada
2	90,600	75,300	1.2	41	indústria	fazenda	diferenciada
3	195,600	52,100	3.75	37	indústria	fazenda	padrão
4	157,800	67,600	2.33	44	indústria	apto.	diferenciada
5	150,800	35,800	4.21	39	autônomo	apto.	padrão
6	133,000	45,300	2.94	29	indústria	fazenda	padrão
7	193,100	73,200	2.64	38	autônomo	casa	diferenciada
8	215,000	77,600	2.77	17	autônomo	fazenda	diferenciada
9	83,000	62,500	1.33	30	autônomo	casa	diferenciada
10	186,100	49,200	3.78	30	indústria	casa	padrão
11	161,500	53,300	3.03	28	autônomo	apto.	diferenciada
12	157,400	63,900	2.46	30	autônomo	fazenda	diferenciada
13	210,000	54,200	3.87	43	autônomo	apto.	diferenciada
14	209,700	53,000	3.96	39	indústria	fazenda	padrão
15	143,200	65,300	2.19	32	indústria	apto.	padrão
16	203,000	64,400	3.15	44	indústria	fazenda	diferenciada
...

```
if Razão Renda-Empréstimo < 1.5 then
    Classe= 'diferenciada'
else if Razão Renda-Empréstimo > 4 then
    Classe= 'padrão'
else if Idade < 40 and Profissão = 'indústria' then
    Classe= 'padrão'
else
    Classe= 'diferenciada'
end if
```

```
if Razão Renda-Empréstimo < 1.5 then
    Classe= 'diferenciada'
else if Razão Renda-Empréstimo > 4 then
    Classe= 'padrão'
else if Idade < 40 and Profissão = 'indústria' then
    Classe= 'padrão'
else
    Classe= 'diferenciada'
end if
```

- ▶ E se fosse possível extrair as regras automaticamente?

Machine Learning

- ▶ É um processo automatizado que extrai padrões de um conjunto de dados para serem aplicados a um único problema do domínio ao qual os dados pertencem
- ▶ Tradução:
 - ▶ É um processo que extrai padrões para atingir **um e apenas um único objetivo**
 - ▶ e.g.: identificação de fraudes em cartão de crédito, sugestão de produtos em loja virtual, ...

- ▶ Quase todos os problemas que se possa imaginar podem ser reduzidos a dois tipos:
- ▶ Predição → predizer um número
 - ▶ probabilidade de que vá chover
 - ▶ dosagem de um certo medicamento a um paciente
 - ▶ ...
- ▶ Classificação → predizer uma categoria
 - ▶ classificar como fraude/não fraude uma transação bancária
 - ▶ classificar como ligar/não ligar para um cliente oferecendo vantagens
 - ▶ classificar qual animal aparece em uma imagem
 - ▶ ...

ID	Profissão	Idade	Salário-Emprestimo	Proporção
				Classe
1	indústria	34	2.96	diferenciada
2	autônomo	41	4.64	padrão
3	autônomo	36	3.22	padrão
4	autônomo	41	3.11	padrão
5	indústria	48	3.80	padrão
6	indústria	61	2.52	diferenciada
7	autônomo	37	1.50	diferenciada
8	autônomo	40	1.93	diferenciada
9	indústria	33	5.25	padrão
10	indústria	32	4.15	padrão

ID	Profissão	Idade	Proporção	Classe
			Salário-Empréstimo	
1	indústria	34	2.96	diferenciada
2	autônomo	41	4.64	padrão
3	autônomo	36	3.22	padrão
4	autônomo	41	3.11	padrão
5	indústria	48	3.80	padrão
6	indústria	61	2.52	diferenciada
7	autônomo	37	1.50	diferenciada
8	autônomo	40	1.93	diferenciada
9	indústria	33	5.25	padrão
10	indústria	32	4.15	padrão

features: conjunto de informações relativos a um caso do domínio

ID	Profissão	Idade	Salário-Emprestimo	Proporção	Classe
1	indústria	34	2.96		diferenciada
2	autônomo	41	4.64		padrão
3	autônomo	36	3.22		padrão
4	autônomo	41	3.11		padrão
5	indústria	48	3.80		padrão
6	indústria	61	2.52		diferenciada
7	autônomo	37	1.50		diferenciada
8	autônomo	40	1.93		diferenciada
9	indústria	33	5.25		padrão
10	indústria	32	4.15		padrão

alvo: resposta correta relativa a um caso do domínio

ID	Profissão	Idade	Salário-Emprestimo	Proporção
				Classe
1	indústria	34	2.96	diferenciada
2	autônomo	41	4.64	padrão
3	autônomo	36	3.22	padrão
4	autônomo	41	3.11	padrão
5	indústria	48	3.80	padrão
6	indústria	61	2.52	diferenciada
7	autônomo	37	1.50	diferenciada
8	autônomo	40	1.93	diferenciada
9	indústria	33	5.25	padrão
10	indústria	32	4.15	padrão

datapoint: um caso do domínio = features + alvo

ID	Profissão	Idade	Salário-Emprestimo	Proporção
				Classe
1	indústria	34	2.96	diferenciada
2	autônomo	41	4.64	padrão
3	autônomo	36	3.22	padrão
4	autônomo	41	3.11	padrão
5	indústria	48	3.80	padrão
6	indústria	61	2.52	diferenciada
7	autônomo	37	1.50	diferenciada
8	autônomo	40	1.93	diferenciada
9	indústria	33	5.25	padrão
10	indústria	32	4.15	padrão

dataset: conjunto de *datapoints*

Definições:

- ▶ **features**: conjunto de informações relativos a um caso do domínio
- ▶ **alvo**: resposta correta relativa a um caso do domínio
- ▶ **datapoint**: um caso do domínio = *features + alvo*
- ▶ **dataset**: conjunto de *datapoints*
- ▶ **algoritmo**: processo usado para aprender as regras
- ▶ **modelo**: conjunto de **regras aprendidas**

- ▶ O algoritmo procura dentre uma série de possíveis modelos o que melhor descreve a relação entre as **features** e o **alvo**
 - ▶ Definem as relações entre **features** e alvo, por isso **precisamos das respostas corretas!**
- ▶ A forma óbvia é procurar modelos que são **consistentes com as datasets**
- ▶ Entretanto, o *machine learning* é “imperfeito” por natureza
 - ▶ Independente do tamanho da *dataset*, eles representam apenas uma porção dos possíveis casos naquele domínio

ID	Filhos	Álcool	Orgânicos	Grupo
1	não	não	não	casal
2	sim	não	sim	família
3	sim	sim	não	família
4	não	não	sim	casal
5	não	sim	sim	solteiro

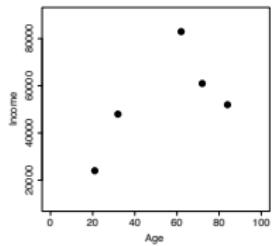
Flh	Álc	Org	Grp	M ₁	M ₂	M ₃	M ₄	M ₅	..
não	não	não	?	casal	casal	solteiro	casal	casal	
não	não	sim	?	solteiro	casal	solteiro	casal	casal	
não	sim	não	?	família	família	solteiro	solteiro	solteiro	
não	sim	sim	?	solteiro	solteiro	solteiro	solteiro	solteiro	
sim	não	não	?	casal	casal	família	família	família	
sim	não	sim	?	casal	família	família	família	família	
sim	sim	não	?	solteiro	família	família	família	família	
sim	sim	sim	?	solteiro	solteiro	família	família	casal	

► 6561 soluções possíveis!

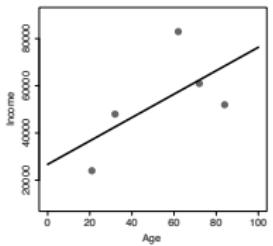
Flh	Álc	Org	Grp	M ₁	M ₂	M ₃	M ₄	M ₅
não	não	não	?	casal	casal	solteiro	casal	casal
não	não	sim	?	solteiro	casal	solteiro	casal	casal
não	sim	não	?	família	família	solteiro	solteiro	solteiro
não	sim	sim	?	solteiro	solteiro	solteiro	solteiro	solteiro
sim	não	não	?	casal	casal	família	família	família
sim	não	sim	?	casal	família	família	família	família
sim	sim	não	?	solteiro	família	família	família	família
sim	sim	sim	?	solteiro	solteiro	família	família	casal

- Só neste slide, 3 soluções consistentes estão visíveis!

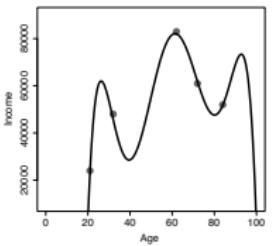
- ▶ Consistência \approx memorização
- ▶ Consistência quando há erros nos dados não é desejável
- ▶ Objetivo: modelo que seja “genérico” e funcione além do *dataset* usado para treinar
 - ▶ Underfitting
 - ▶ Overfitting



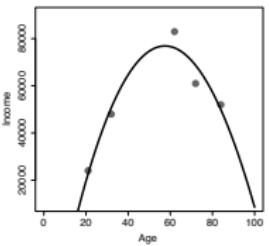
(a) Features



(b) Underfitting



(c) Overfitting



(d) Correto

4 tipos básicos de algoritmos em *machine learning*

- ▶ baseados em informação
- ▶ baseados em similaridade
- ▶ baseados em probabilidade
- ▶ baseados em erro

Métodos baseados em informação



(a) Brian



(b) John



(c) Aphra



(d) Aoife

Homem	Cabelo Longo	Óculos	Nome
sim	não	sim	Brian
sim	não	não	John
não	sim	não	Aphra
não	não	não	Aoife



(a) Brian



(b) John



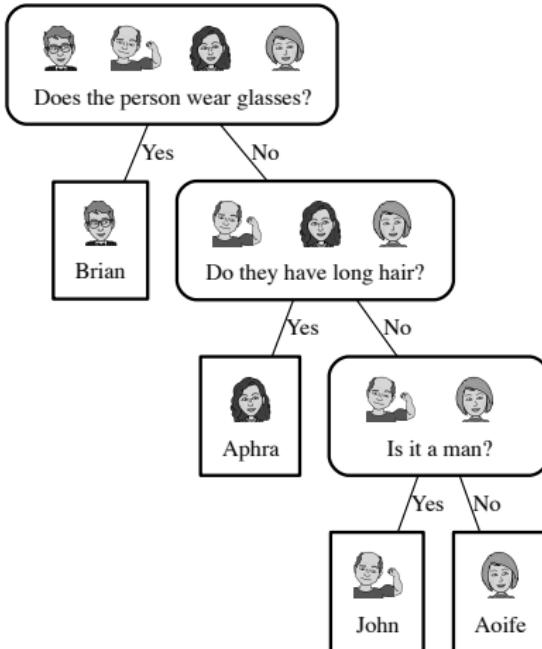
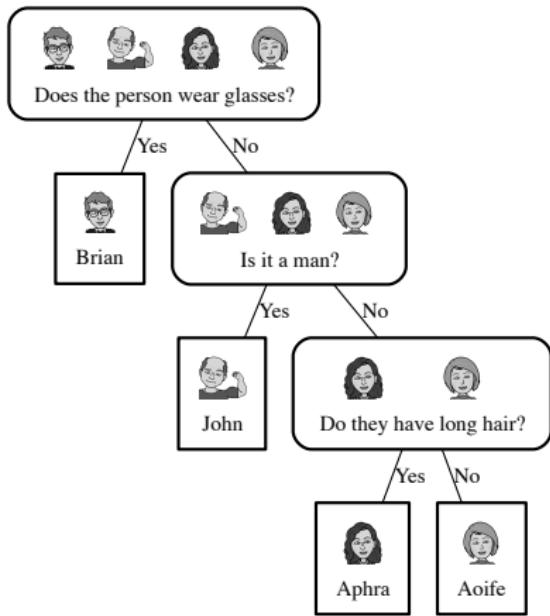
(c) Aphra



(d) Aoife

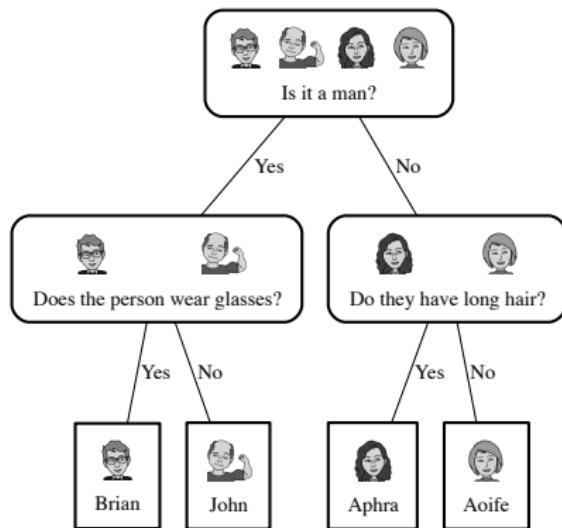
► Qual pergunta você faria primeiro?

- A. Essa pessoa usa óculos?
- B. Essa pessoa é um homem?



- ▶ Em ambos os casos
 - ▶ um caminho tem 1 pergunta
 - ▶ um caminho tem 2 perguntas
 - ▶ um caminho tem 3 perguntas
- ▶ Portanto, se começarmos com a pergunta **A.**, temos em média

$$\frac{1 + 2 + 3 + 3}{4} = 2.25 \text{ perguntas}$$



- ▶ Neste caso, todos os caminhos tem 2 perguntas
- ▶ Portanto, se começarmos com a pergunta **B.**, temos em média

$$\frac{2 + 2 + 2 + 2}{4} = 2 \text{ perguntas}$$

- ▶ A ideia por trás destes métodos é descobrir quais *features* carregam mais informação na hora de dividir os alvos em grupos homogêneos
- ▶ quando dividimos as *datasets*, precisamos saber qual a “pureza” dos conjuntos resultantes
 - ▶ **entropia**
 - ▶ ou quantos *bits* nós precisamos pra representar a mesma informação
- ▶ após a divisão, temos que saber o quanto ganhamos em termos da entropia nos conjuntos resultantes em relação a entropia inicial
 - ▶ **information gain**
 - ▶ medida da redução de entropia quando se escolhe uma determinada *feature*

► Entropia

$$H(t, \mathcal{D}) = - \sum_{l \in levels(t)} (P(t = l) \times \log_2(P(t = l)))$$

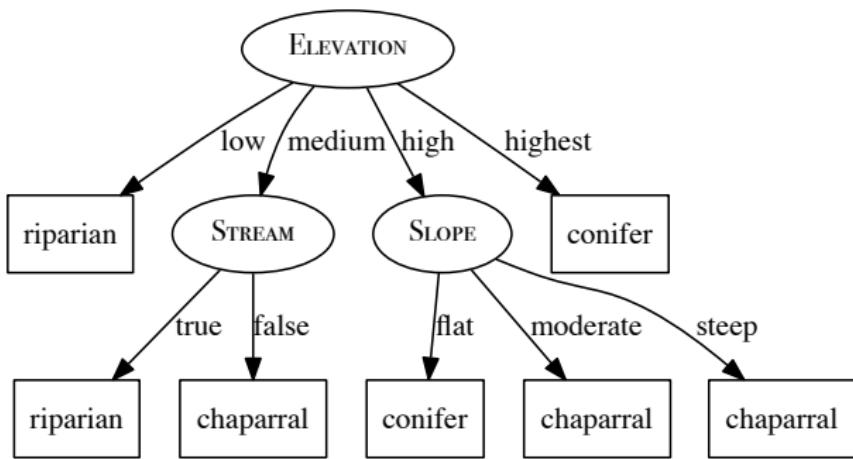
► Information gain

$$rem(d, \mathcal{D}) = \sum_{l \in levels(d)} \underbrace{\frac{|\mathcal{D}_{d=l}|}{|\mathcal{D}|}}_{\text{"peso"}} \times \underbrace{H(t, \mathcal{D}_{d=l})}_{\substack{\text{entropia do} \\ \text{conjunto } \mathcal{D}_{d=l}}}$$

$$IG(d, \mathcal{D}) = H(t, \mathcal{D}) - rem(d, \mathcal{D})$$

ID	Stream	Slope	Elevation	Vegetation
1	false	steep	high	chaparral
2	true	moderate	low	riparian
3	true	steep	medium	riparian
4	false	steep	medium	chaparral
5	false	flat	high	conifer
6	true	steep	highest	conifer
7	true	steep	high	chaparral

Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>



Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

Métodos baseados em similaridade

- ▶ O ano é 1798 e você é o Tenente-Coronel David Collins do navio HMS Calcutta, explorando a região da Nova Gales do Sul, Austrália
- ▶ Após uma expedição, os seus soldados reportam que encontraram um animal estranho próximo de um rio
- ▶ Você pede uma descrição do animal e os soldados respondem que não conseguiram ver o animal com clareza, mas eles repararam que ele possuia bico e pés de pato e que ele rosnou para reles
- ▶ Para preparar a expedição do dia seguinte, você precisa determinar se aquele animal é uma ameaça ou não para os seus soldados

	Grrrh!	↙	↗	Score
	✓	✗	✗	1
	✗	✓	✗	1
	✗	✓	✓	2

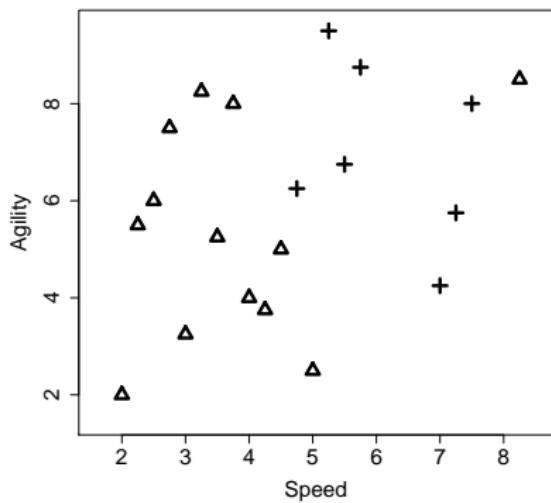
- ▶ Métodos baseados em similaridade utilizam 2 conceitos fundamentais
 - ▶ Espaço de *features* (ou *feature space*)
 - ▶ Medida de distância

ID	Speed	Agility	Draft
1	2.50	6.00	No
2	3.75	8.00	No
3	2.25	5.50	No
4	3.25	8.25	No
5	2.75	7.50	No
6	4.50	5.00	No
7	3.50	5.25	No
8	3.00	3.25	No
9	4.00	4.00	No
10	4.25	3.75	No

ID	Speed	Agility	Draft
11	2.00	2.00	No
12	5.00	2.50	No
13	8.25	8.50	No
14	5.75	8.75	Yes
15	4.75	6.25	Yes
16	5.50	6.75	Yes
17	5.25	9.50	Yes
18	7.00	4.25	Yes
19	7.50	8.00	Yes
20	7.25	5.75	Yes

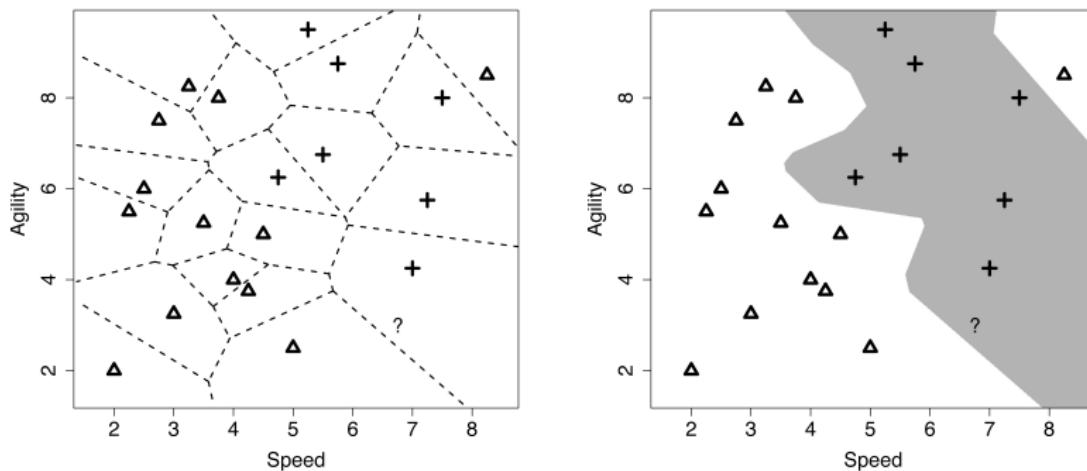
Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

- ▶ *feature space* é um espaço abstrato de N dimensões, uma para cada *feature*



Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

- ▶ Uma medida de similaridade mede a “distância” entre *datapoints* no *feature space*

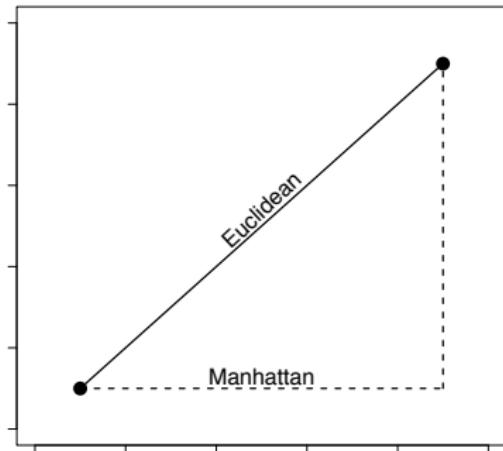


Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

$$\text{Manhattan}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^m \text{abs}(\mathbf{a}[i] - \mathbf{b}[i])$$

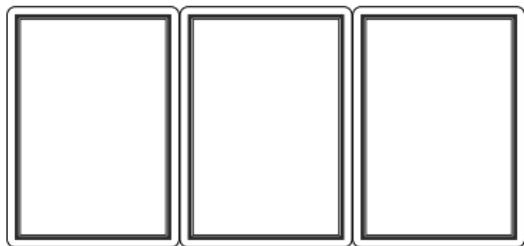
$$\text{Euclidean}(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^m (\mathbf{a}[i] - \mathbf{b}[i])^2}$$

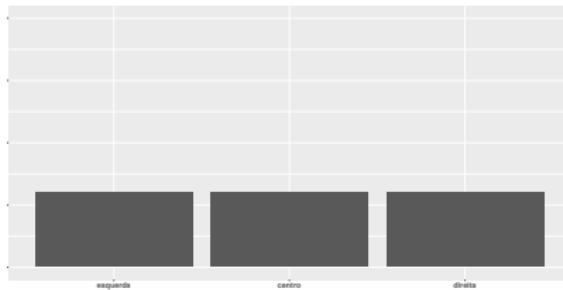
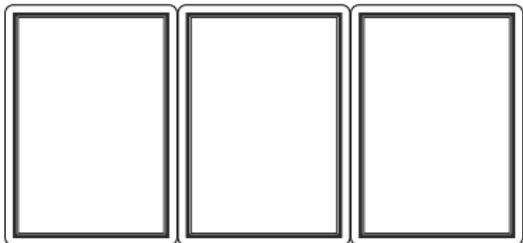
$$\text{Minkowski}(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^m \text{abs}(\mathbf{a}[i] - \mathbf{b}[i])^p \right)^{\frac{1}{p}}$$

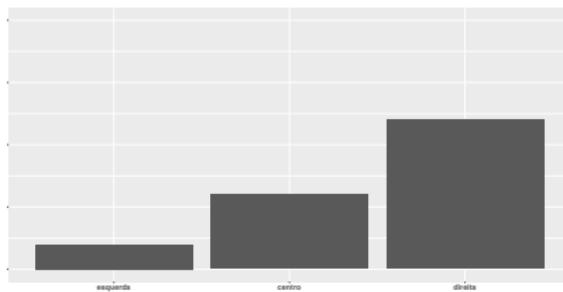
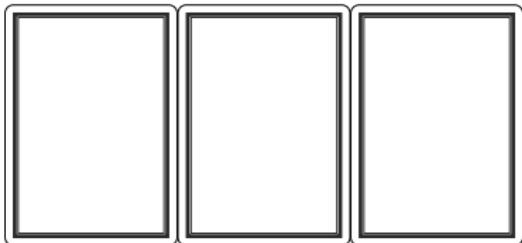


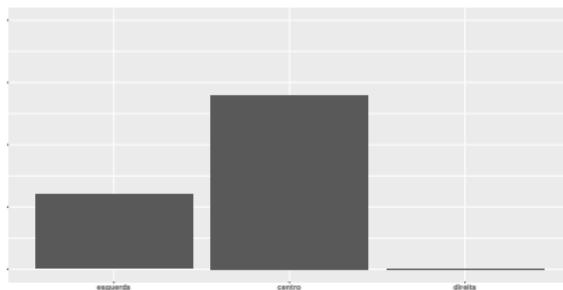
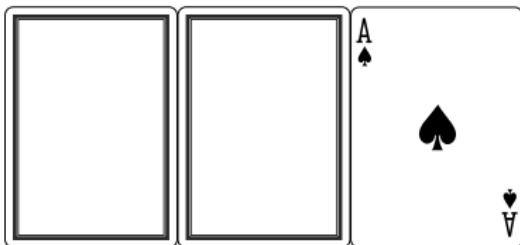
Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

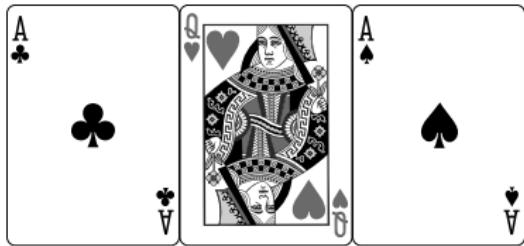
Métodos baseados em probabilidade











- ▶ Nestes métodos, nós utilizamos a probabilidade de um evento (no caso o valor de uma *feature*) para determinar qual a melhor escolha
- ▶ Sempre que temos novos dados em mãos, nós revisamos as probabilidades e fazemos ajustes
- ▶ o algoritmo cria “fatores” baseados nas *features* e no alvo e os utiliza para calcular a probabilidade de cada tipo de alvo

$$\mathbb{M}(\mathbf{q}) = \operatorname{argmax}_{l \in \text{alvos}} \left(\prod_{i=1}^m P(\mathbf{q}[i] \mid a = l) \right) \times P(a = l)$$

ID	CH	GC	ACC	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

$P(fr)$	=	0.3	$P(\neg fr)$	=	0.7
$P(CH = 'none' fr)$	=	0.1666	$P(CH = 'none' \neg fr)$	=	0
$P(CH = 'paid' fr)$	=	0.1666	$P(CH = 'paid' \neg fr)$	=	0.2857
$P(CH = 'current' fr)$	=	0.5	$P(CH = 'current' \neg fr)$	=	0.2857
$P(CH = 'arrears' fr)$	=	0.1666	$P(CH = 'arrears' \neg fr)$	=	0.4286
$P(GC = 'none' fr)$	=	0.8334	$P(GC = 'none' \neg fr)$	=	0.8571
$P(GC = 'guarantor' fr)$	=	0.1666	$P(GC = 'guarantor' \neg fr)$	=	0
$P(GC = 'coapplicant' fr)$	=	0	$P(GC = 'coapplicant' \neg fr)$	=	0.1429
$P(ACC = 'own' fr)$	=	0.6666	$P(ACC = 'own' \neg fr)$	=	0.7857
$P(ACC = 'rent' fr)$	=	0.3333	$P(ACC = 'rent' \neg fr)$	=	0.1429
$P(ACC = 'free' fr)$	=	0	$P(ACC = 'free' \neg fr)$	=	0.0714

Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

$P(fr)$	=	0.3	$P(\neg fr)$	=	0.7
$P(\text{CH} = \text{'none'} fr)$	=	0.1666	$P(\text{CH} = \text{'none'} \neg fr)$	=	0
$P(\text{CH} = \text{'paid'} fr)$	=	0.1666	$P(\text{CH} = \text{'paid'} \neg fr)$	=	0.2857
$P(\text{CH} = \text{'current'} fr)$	=	0.5	$P(\text{CH} = \text{'current'} \neg fr)$	=	0.2857
$P(\text{CH} = \text{'arrears'} fr)$	=	0.1666	$P(\text{CH} = \text{'arrears'} \neg fr)$	=	0.4286
$P(\text{GC} = \text{'none'} fr)$	=	0.8334	$P(\text{GC} = \text{'none'} \neg fr)$	=	0.8571
$P(\text{GC} = \text{'guarantor'} fr)$	=	0.1666	$P(\text{GC} = \text{'guarantor'} \neg fr)$	=	0
$P(\text{GC} = \text{'coapplicant'} fr)$	=	0	$P(\text{GC} = \text{'coapplicant'} \neg fr)$	=	0.1429
$P(\text{ACC} = \text{'own'} fr)$	=	0.6666	$P(\text{ACC} = \text{'own'} \neg fr)$	=	0.7857
$P(\text{ACC} = \text{'rent'} fr)$	=	0.3333	$P(\text{ACC} = \text{'rent'} \neg fr)$	=	0.1429
$P(\text{ACC} = \text{'free'} fr)$	=	0	$P(\text{ACC} = \text{'free'} \neg fr)$	=	0.0714

CH	GC	ACC	Fraud
paid	none	rent	?

Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

$P(fr)$	=	0.3	$P(\neg fr)$	=	0.7
$P(CH = 'paid' fr)$	=	0.1666	$P(CH = 'paid' \neg fr)$	=	0.2857
$P(GC = 'none' fr)$	=	0.8334	$P(GC = 'none' \neg fr)$	=	0.8571
$P(ACC = 'rent' fr)$	=	0.3333	$P(ACC = 'rent' \neg fr)$	=	0.1429
$\left(\prod_{k=1}^m P(\mathbf{q}[k] fr) \right) \times P(fr) = 0.0139$					
$\left(\prod_{k=1}^m P(\mathbf{q}[k] \neg fr) \right) \times P(\neg fr) = 0.0245$					

CH	GC	ACC	Fraud
paid	none	rent	False

Fonte: John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Work Examples and Case Studies*. MIT Press.
<http://machinelearningbook.com/>

Métodos baseados em erro

- ▶ Dado uma *dataset* \mathcal{D} com N *datapoints*: $(\mathbf{x}_n, y_n) \in (\mathcal{X}, \mathcal{Y})$
 - ▶ onde \mathbf{x}_n são as *features* e y_n é o alvo, ambos relativos a um *datapoint* n
- ▶ algoritmo itera sobre o *dataset* e “aprende” uma função parametrizada $f : \mathcal{X} \rightarrow \mathcal{Y}$
 - ▶ esta função descreve a relação entre as *features* e o alvo
 - ▶ parâmetros também são chamados de “pesos” e controlam a saída retornada pela função
 - ▶ modelo é formado pela função e os parâmetros
- ▶ Para aprender f , o algoritmo utiliza uma medida de performance $\mathcal{L}(y_n, f(\mathbf{x}_n))$
 - ▶ Aprendizado do modelo neste caso também é chamado de *otimização*

$$z = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n$$

$$= w_0 + \sum_{i=1}^n w_i x_i$$

$$\mathcal{M}_w(\mathbf{x}) = \sigma(z) \text{ ou } \mathcal{M}_w(\mathbf{x}) = z$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

- ▶ Medida de performance:

$$\mathcal{L}_2(\mathcal{M}_w, \mathcal{D}) = \frac{1}{N} \sum_{j=1}^N [y_j - \mathcal{M}_w(x_j)]^2$$

- ▶ Se criarmos um gráfico das possíveis combinações de parâmetros, teremos o *espaço de parâmetros* (ou *weight space*)

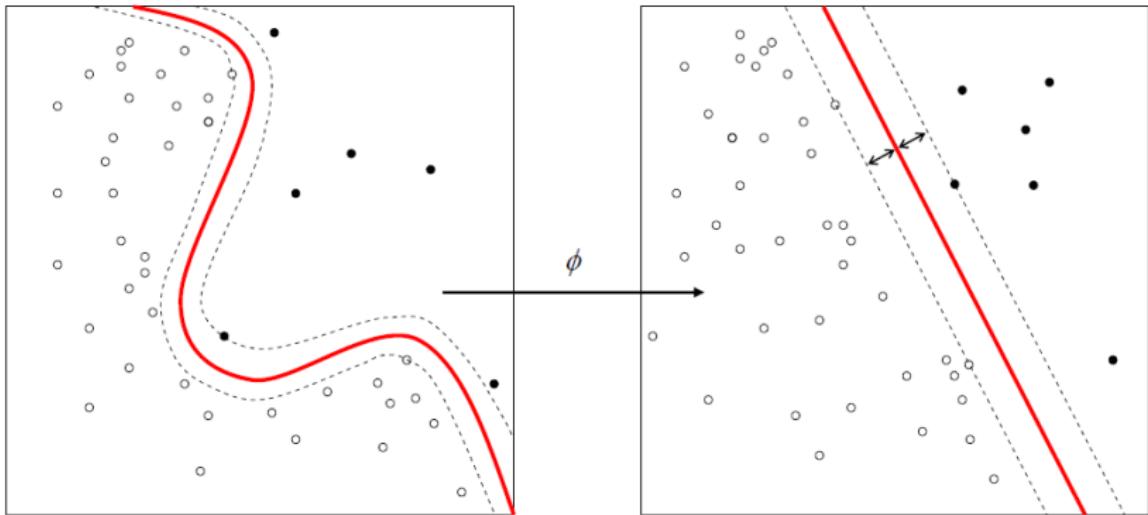
- ▶ Um dos problemas com este algoritmo é o fato dele ser muito bom quando os *datapoints* são linearmente separáveis
 - ▶ Difícilmente os problemas que estamos tentando resolver com IA são linearmente separáveis!
- ▶ Um dos métodos baseados em erro de maior sucesso e que lida com *datapoints* que não são linearmente separáveis é o Support Vector Machine (SVM)
 - ▶ SVM utiliza funções especiais chamadas *kerneis* que vão projetar as *features* em um espaço no qual os *datapoints* serão linearmente separáveis!
 - ▶ *kerneis* são o “motor” das redes neurais artificiais

- ▶ kerneis mais comuns:

$$\text{linear kernel}(\mathbf{d}, \mathbf{q}) = \mathbf{d} \cdot \mathbf{q} + c$$

$$\text{polynomial kernel}(\mathbf{d}, \mathbf{q}) = (\mathbf{d} \cdot \mathbf{q} + 1)^p$$

$$\text{Gaussian kernel}(\mathbf{d}, \mathbf{q}) = \exp(-\gamma \|\mathbf{d} - \mathbf{q}\|^2)$$



Fonte: <http://digitheadsblognotebook.blogspot.ie/2011/11/support-vector-machines.html>

Resumo

► Modelos paramétricos vs. não-paramétricos

- ▶ paramétrico → tamanho do modelo é independente do tamanho do *dataset*
 - ▶ exemplo: métodos baseados em erro
- ▶ não-paramétrico → tamanho do modelo cresce com o tamanho do *dataset*
 - ▶ exemplo: métodos baseados em informação

► Modelos gerativos vs. discriminativos

- ▶ gerativos → aprendem as características da distribuição que gerou o *dataset* e podem ser usados para gerar *datapoints* com as mesmas características
 - ▶ exemplos: métodos baseados em similaridade e em probabilidade
- ▶ discriminativos → aprende o limite que divide 2 ou mais categorias de alvos
 - ▶ exemplo: métodos baseados em erro e em informação

► O que tentar primeiro:

- Alvo é numérico → modelos baseados em erro
- Alvo é categoria → modelos baseados em informação ou probabilidade
- Maioria das *features* é numérica e
 - o alvo numérico → modelos baseados em erro
 - o alvo categoria → modelos baseados em similaridade
- Maioria das *features* é categoria → modelos baseados em informação ou probabilidade

Laboratório

Introdução ao Machine Learning com Scikit Learn