Introdução a Ciencia de Dados

Hitoshi Nagano, Ph.D.

FGV MANAGEMENT

Redes Neurais Artificiais

FGV MANAGEMENT

roteiro

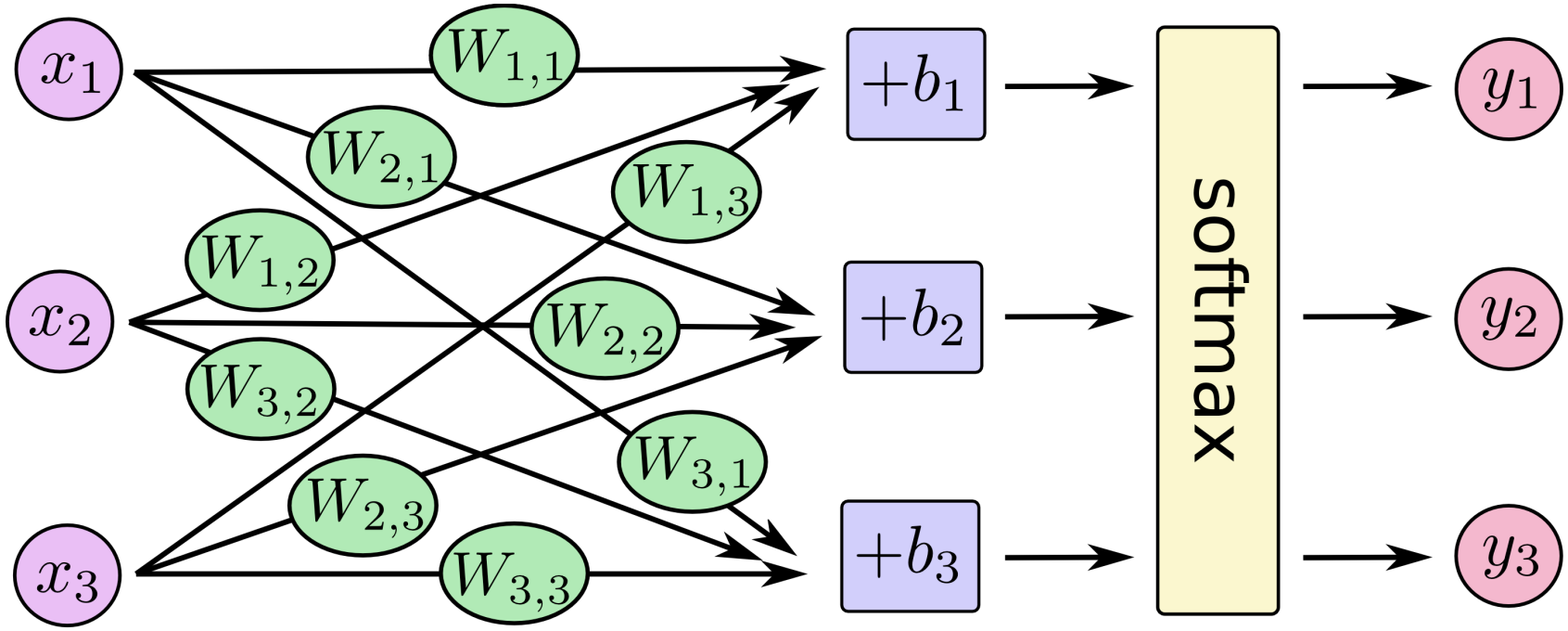
FGV MANAGEMENT

- arquitetura
- forward propagation
- otimização
- backpropagation
- tensorflow & keras

referencias

FGV MANAGEMENT

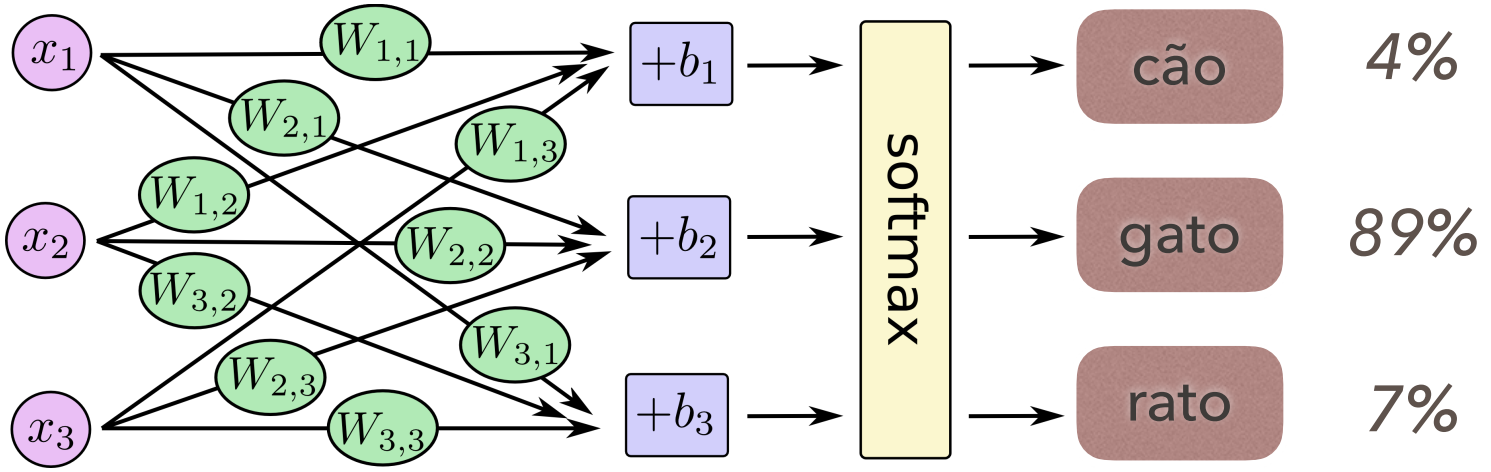
- livros
  http://neuralnetworksanddeeplearning.com
  http://www.deeplearningbook.org/lecture_slides.html
- cursos
  http://cs231n.github.io/
  https://web.stanford.edu/class/cs224n/
- blogs
  http://colah.github.io/

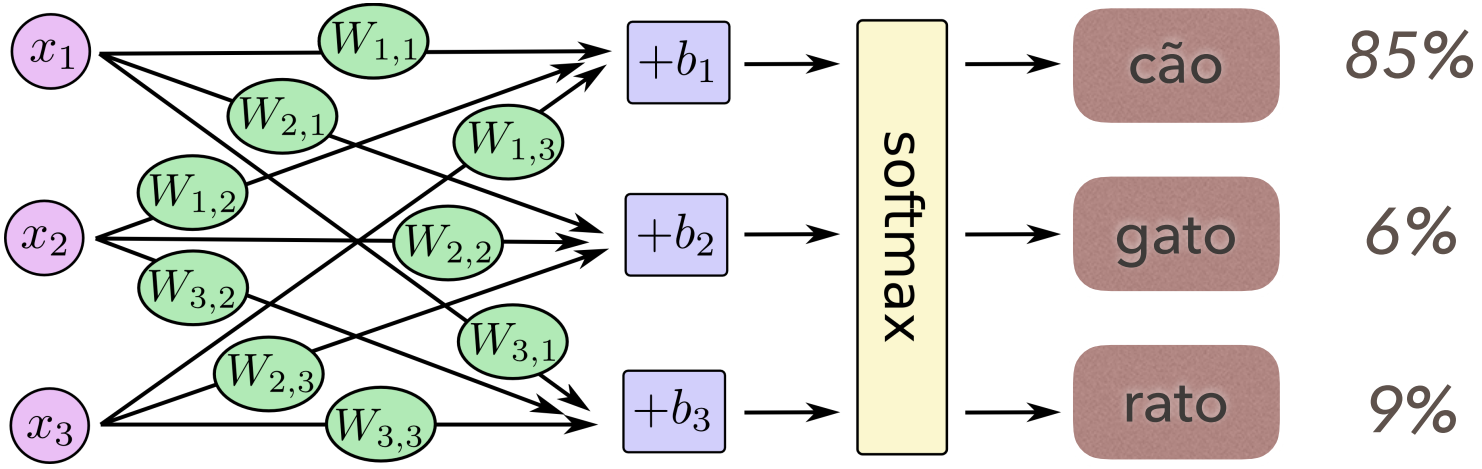


https://www.tensorflow.org/get_started/mnist/beginners

qual é o problema?

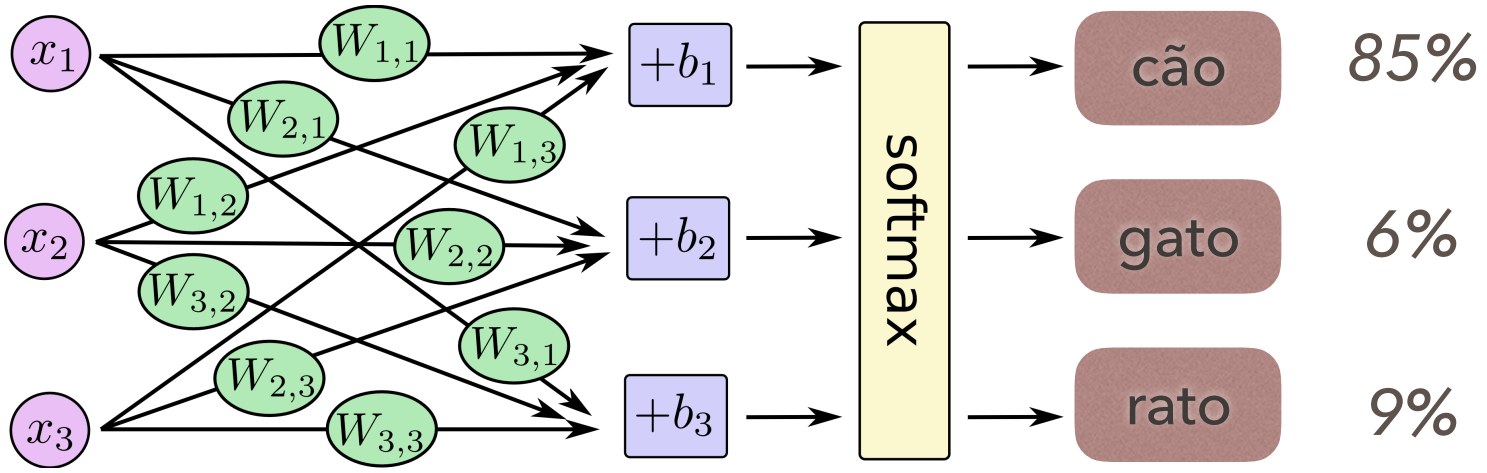
FGV MANAGEMENT





qual é o problema?

FGV MANAGEMENT

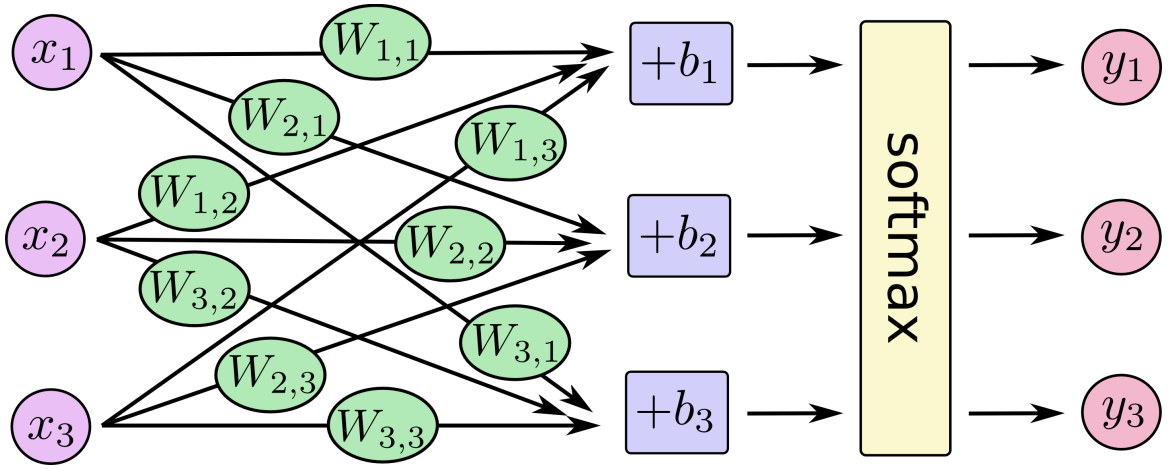


x

	pixel1	pixel2	pixel3	...	pixel n
🐱	2	133	25	...	77
🐶	3	3	55	...	89
🐭	254	255	253	...	255
😼	8	8	10	...	21

y

cão	gato	rato
0	1	0
1	0	0
0	0	1
0	1	0

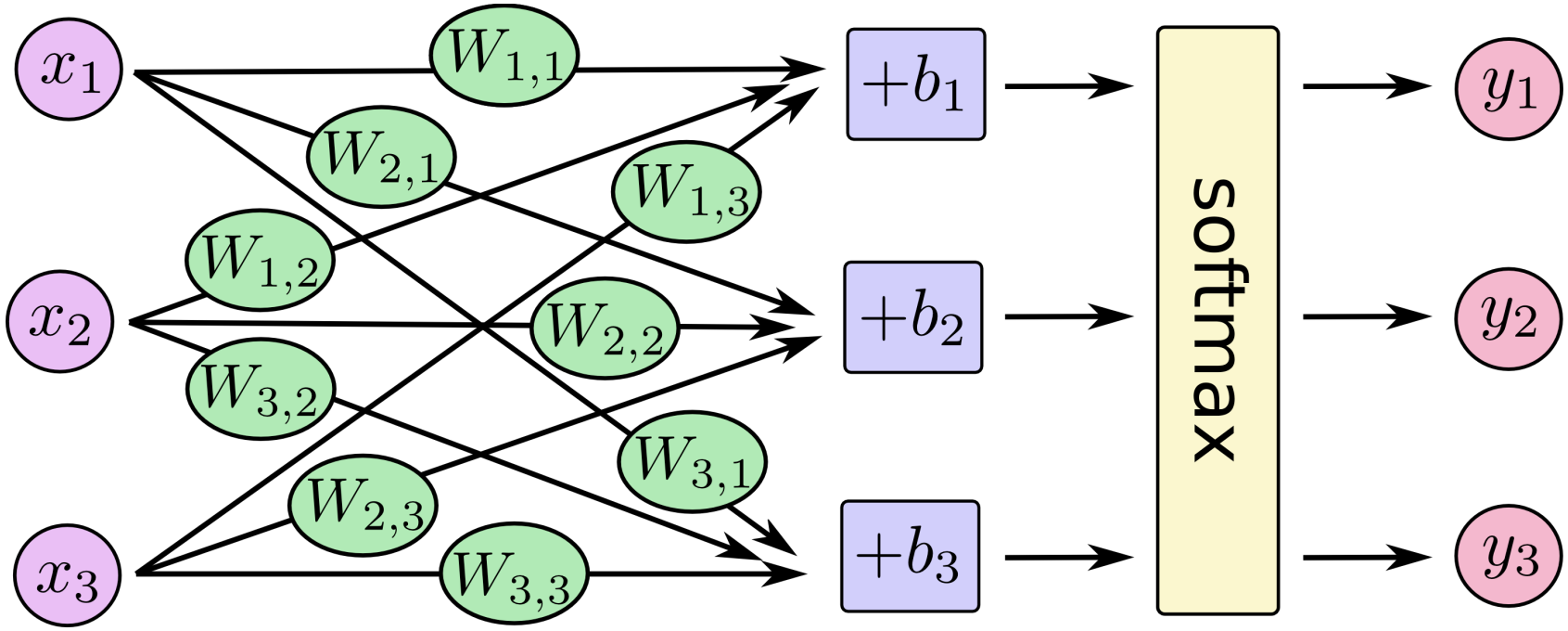


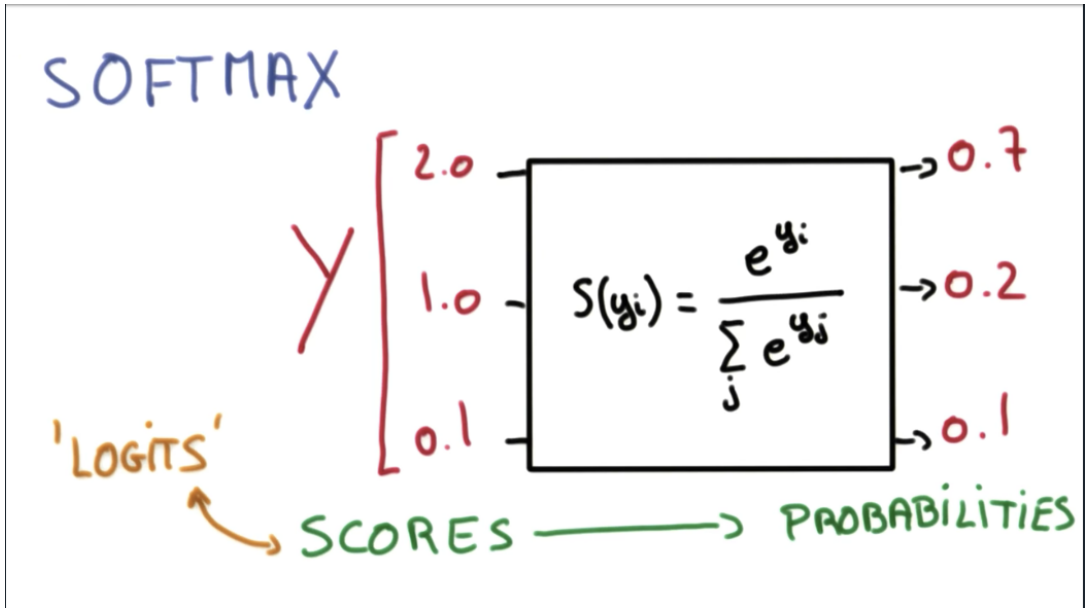
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{softmax}\left(\begin{bmatrix} W_{1,1} & W_{1,2} & W_{1,3} \\ W_{2,1} & W_{2,2} & W_{2,3} \\ W_{3,1} & W_{3,2} & W_{3,3} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right)$$

$$y = softmax(Wx + b)$$

E... o que é esse softmax?

FGV MANAGEMENT

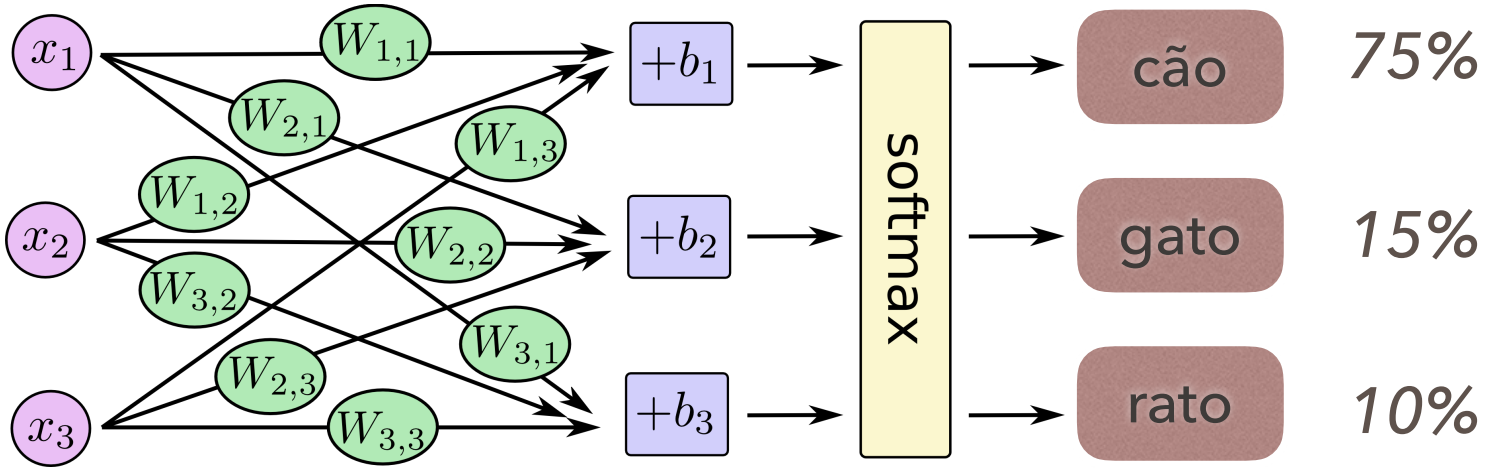




Fonte: Udacity

	scores	e^scores	probabilidades
y_0	2	7.4	0.7
y_1	1	2.7	0.2
y_2	0.1	1.1	0.1
fator de normalização ➜		11.2	

ESSE MODELO É “BOM”?



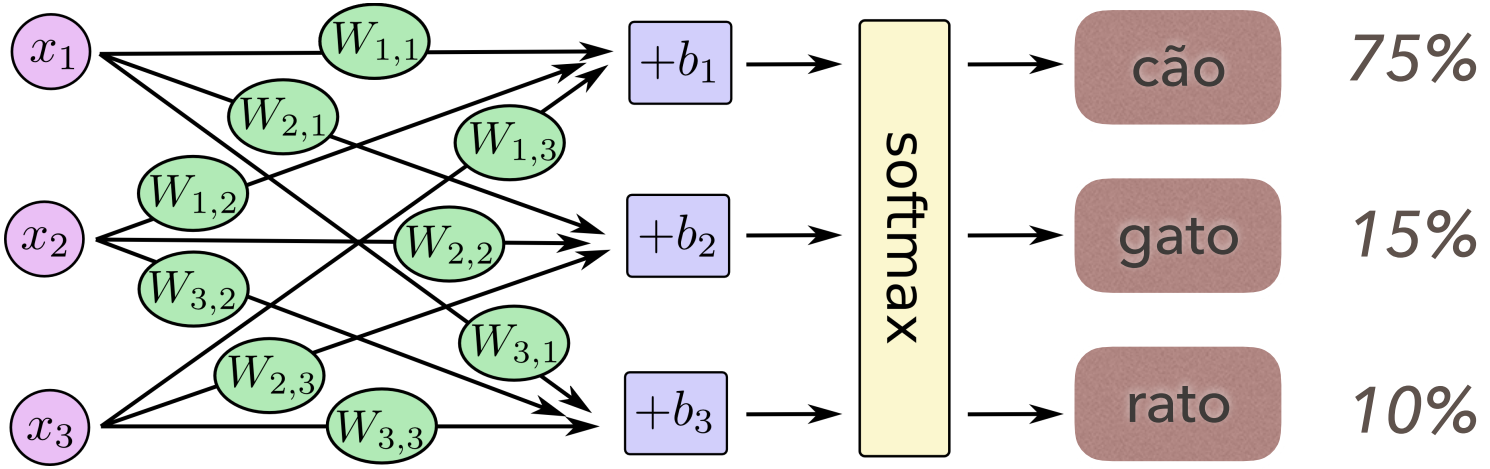
erra menos
acerta mais

→ *“bons” W & b* →

“bom” modelo

como achar "bons" pesos?

FGV MANAGEMENT



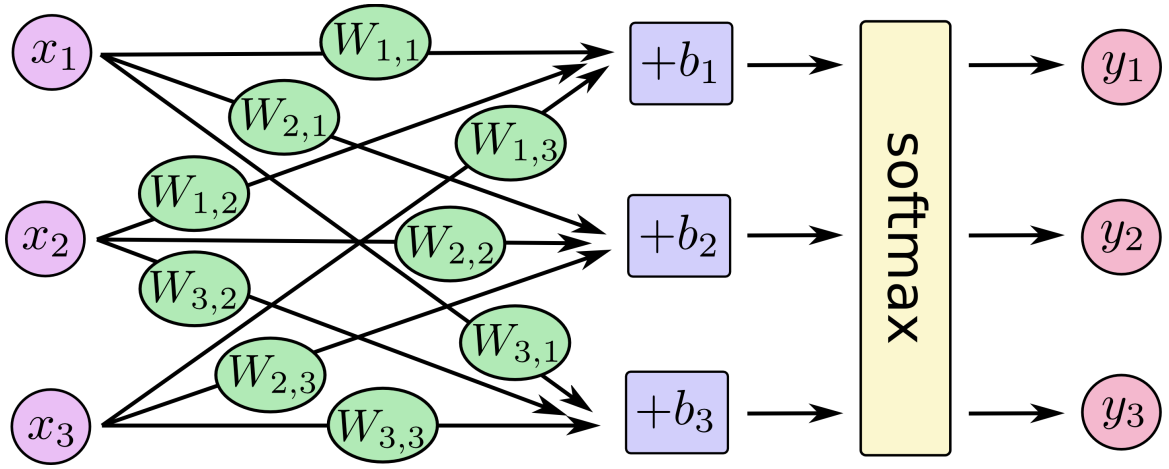
GRADIENT DESCENT

HTTP://WWW.KDNUGGETS.COM/2017/04/SIMPLE-UNDERSTAND-GRADIENT-DESCENT-ALGORITHM.HTML

https://docs.google.com/spreadsheets/d/1TFHcFix5zN5ikqFkJC6mlXg83LxGJAeVJl3BddvYsPM/edit?usp=sharing

como quantificar "BOM"???

FGV MANAGEMENT



x

	pixel1	pixel2	pixel3	...	pixel n
🐱	2	133	25	...	77
🐶	3	3	55	...	89
🐭	254	255	253	...	255
😺	8	8	10	...	21

y

cão	gato	rato
0	1	0
1	0	0
0	0	1
0	1	0

y_pred

cão	gato	rato
0.06	0.85	0.09
0.77	0.20	0.03
0.32	0.15	0.53
0.40	0.20	0.30

como quantificar "BOM"???

FGV MANAGEMENT

	x					*y*			*y_pred*		
	pixel1	pixel2	pixel3	...	pixel n	cão	gato	rato	cão	gato	rato
	2	133	25	...	77	0	1	0	0.06	0.85	0.09
	3	3	55	...	89	1	0	0	0.77	0.20	0.03
	254	255	253	...	255	0	0	1	0.32	0.15	0.53
	8	8	10	...	21	0	1	0	0.40	0.20	0.30

CROSS-ENTROPY

cross-entropy ou log-loss

TRUE PREDS

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j}\log(p_{i,j})$$

https://www.kaggle.com/wiki/LogLoss

cross-entropy ou log-loss

FGV MANAGEMENT

PREDS

	label 0	label 1	label 2	label 3
cliente 0	0.069	0.213	0.698	0.020
cliente 1	0.164	0.426	0.081	0.329
cliente 2	0.278	0.370	0.269	0.083
cliente 3	0.055	0.405	0.460	0.080
cliente 4	0.124	0.666	0.145	0.066
cliente 5	0.569	0.045	0.214	0.172
cliente 6	0.421	0.091	0.353	0.135
cliente 7	0.215	0.323	0.279	0.182
cliente 8	0.139	0.123	0.347	0.390
cliente 9	0.155	0.036	0.311	0.498

TRUE

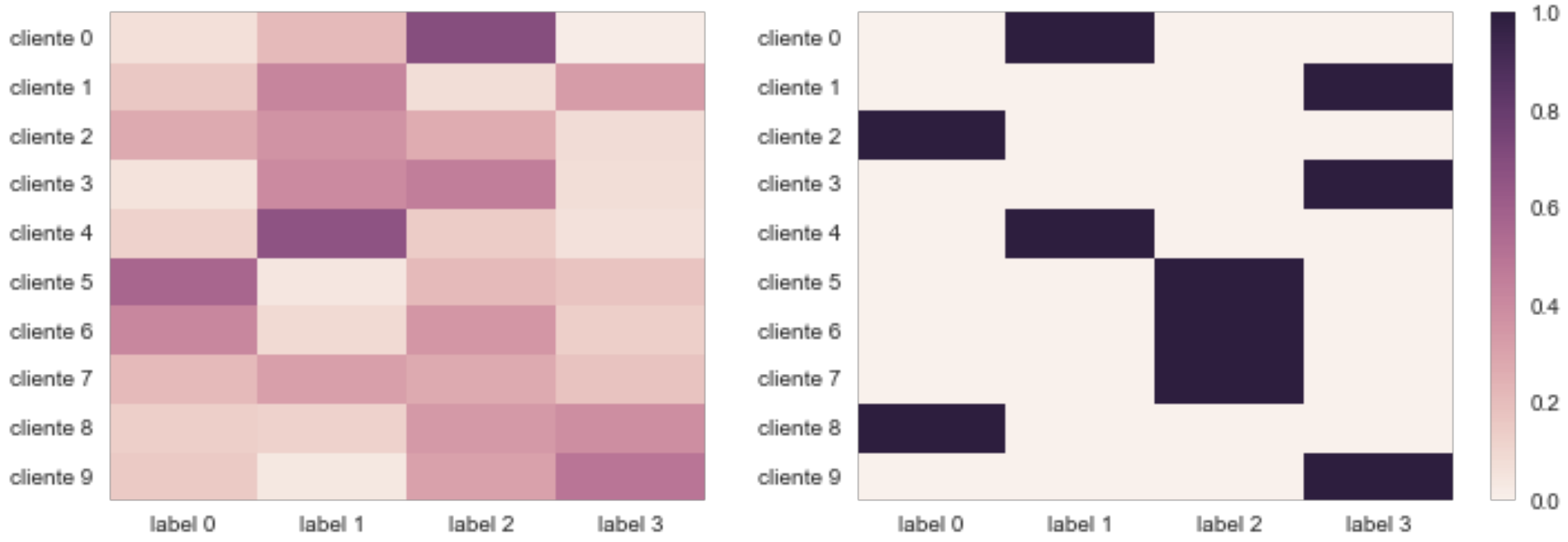
	label 0	label 1	label 2	label 3
cliente 0	0	1	0	0
cliente 1	0	0	0	1
cliente 2	1	0	0	0
cliente 3	0	0	0	1
cliente 4	0	1	0	0
cliente 5	0	0	1	0
cliente 6	0	0	1	0
cliente 7	0	0	1	0
cliente 8	1	0	0	0
cliente 9	0	0	0	1

cross-entropy ou log-loss

FGV MANAGEMENT

PREDS

TRUE

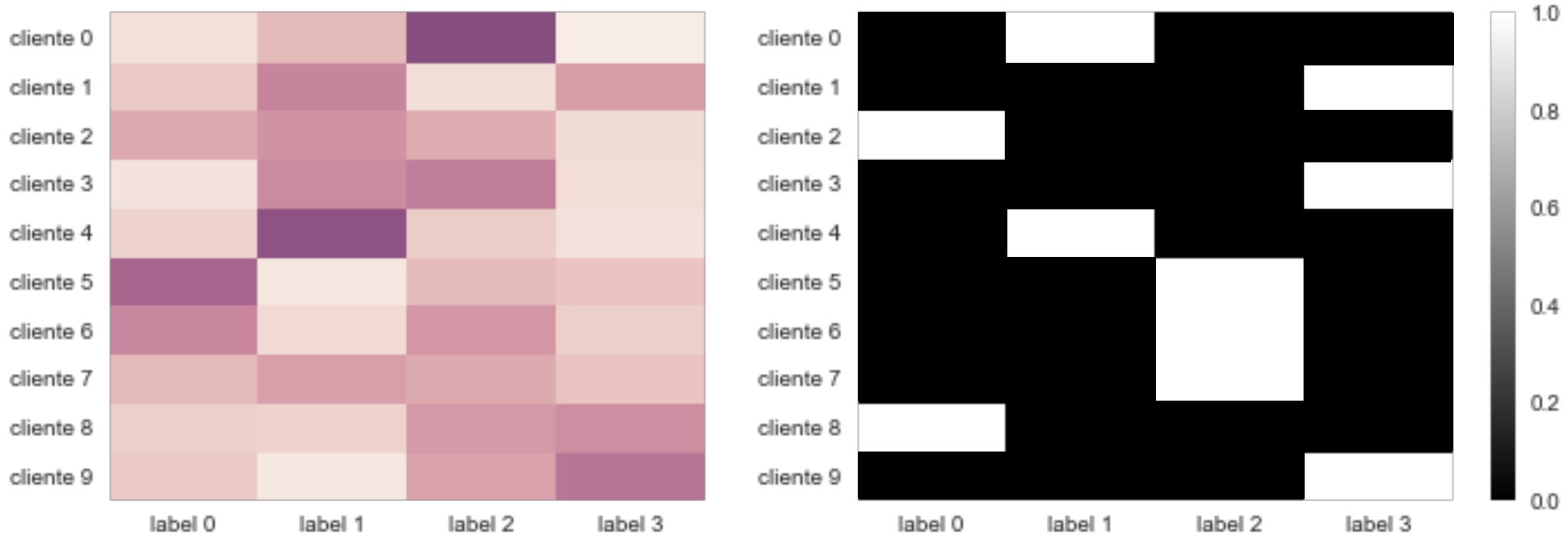


cross-entropy ou log-loss

FGV MANAGEMENT

PREDS

TRUE



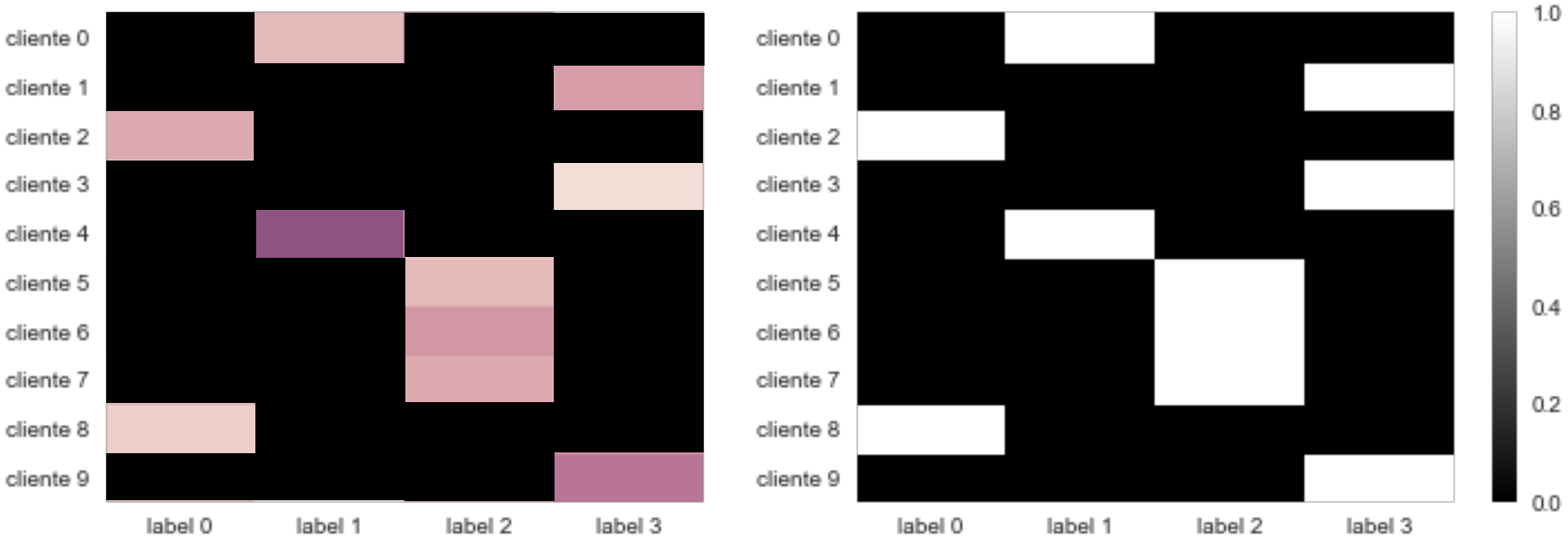
$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j}\log(p_{i,j})$$

cross-entropy ou log-loss

FGV MANAGEMENT

PREDS

TRUE



$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j}\log(p_{i,j})$$

como quantificar "BOM"???

FGV MANAGEMENT

y

cão	gato	rato
0	1	0
1	0	0
0	0	1
0	1	0

y_pred

cão	gato	rato
0.06	0.85	0.09
0.77	0.20	0.03
0.32	0.15	0.53
0.40	0.20	0.40

CROSS-ENTROPY MENOR

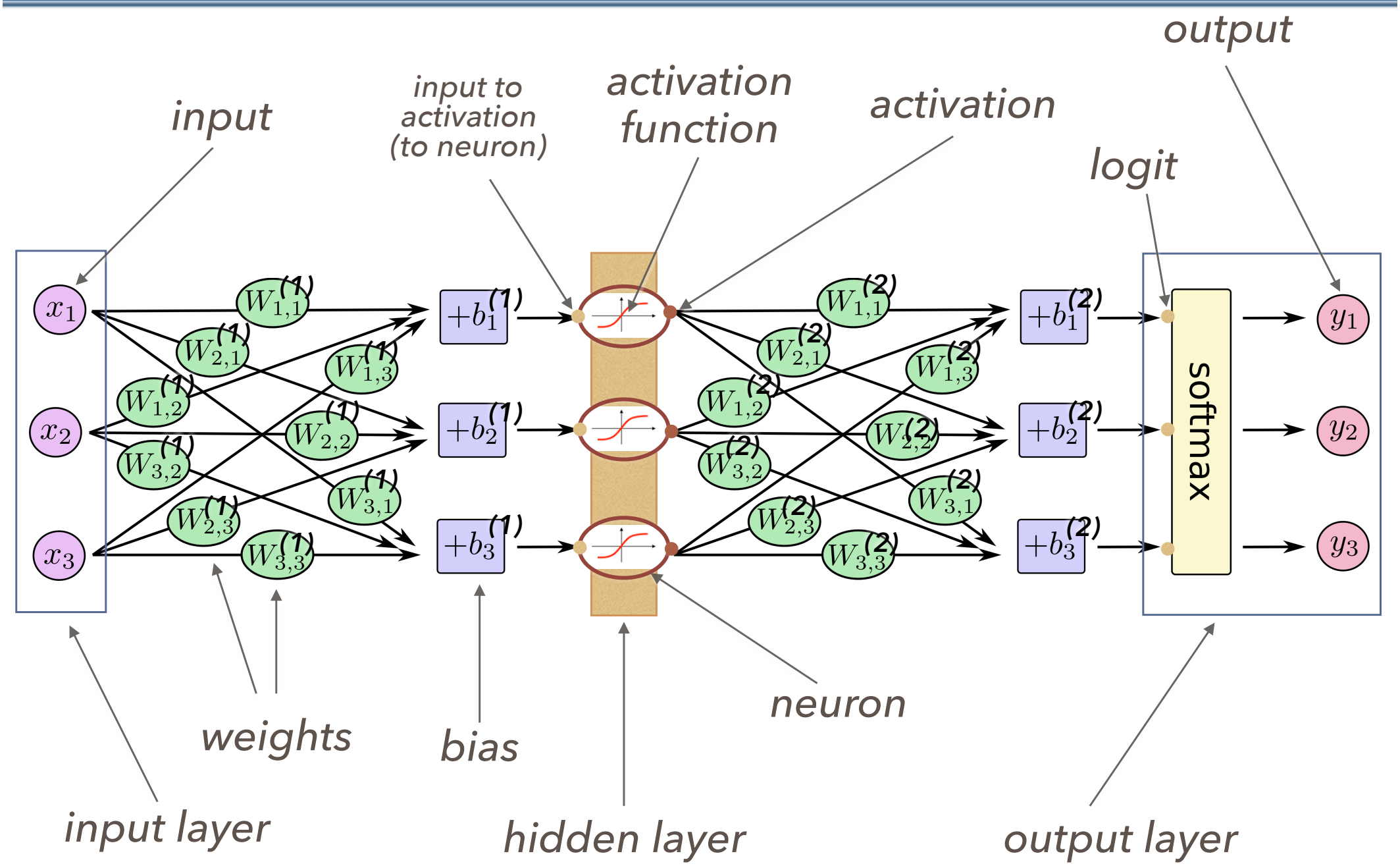
y

cão	gato	rato
0	1	0
1	0	0
0	0	1
0	1	0

y_pred

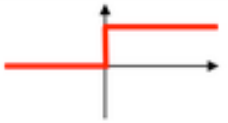
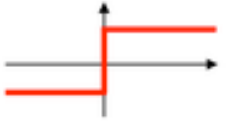
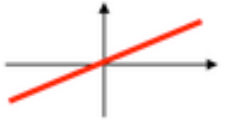
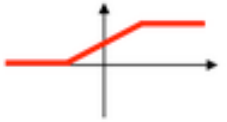
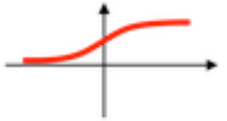
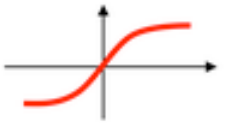
cão	gato	rato
0.60	0.15	0.25
0.24	0.20	0.56
0.32	0.03	0.65
0.40	0.15	0.45

CROSS-ENTROPY MAIOR

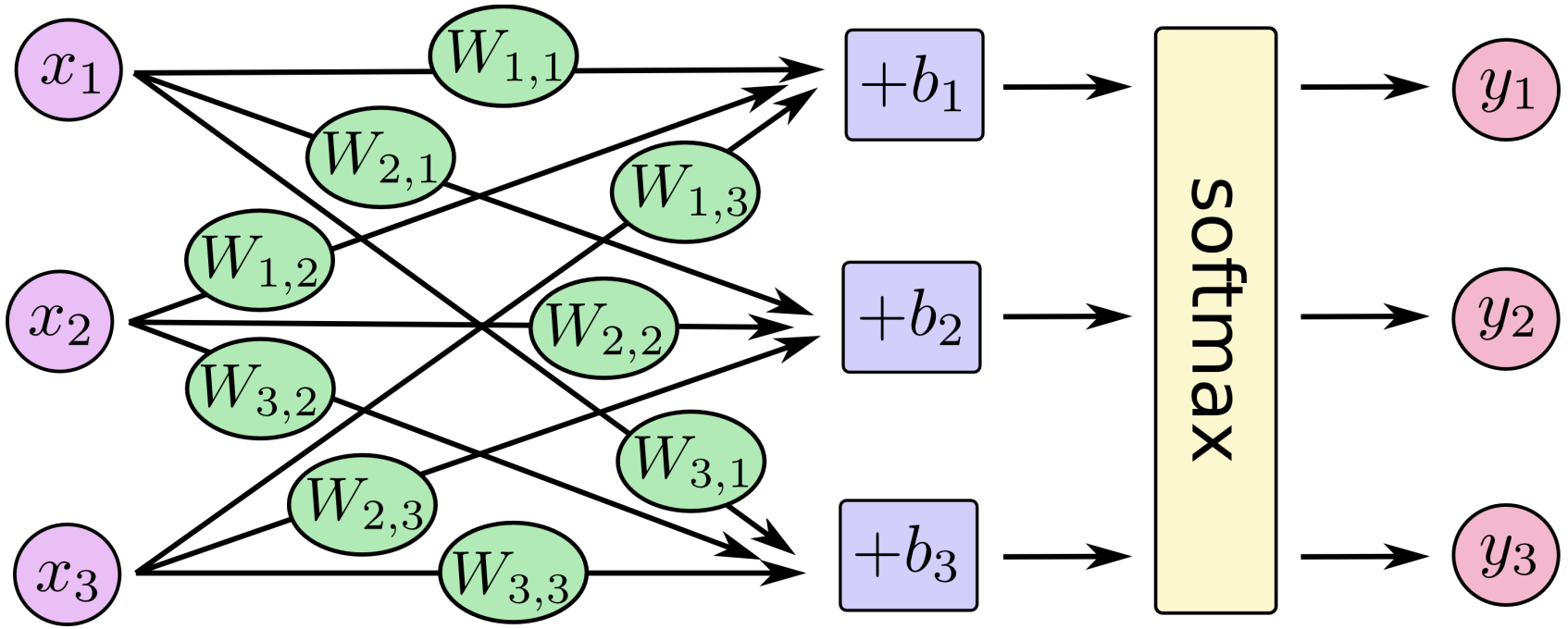

output
input
input to activation (to neuron)
activation function
activation
logit
softmax
weights
bias
neuron
input layer
hidden layer
output layer

diferentes ativações

FGV MANAGEMENT

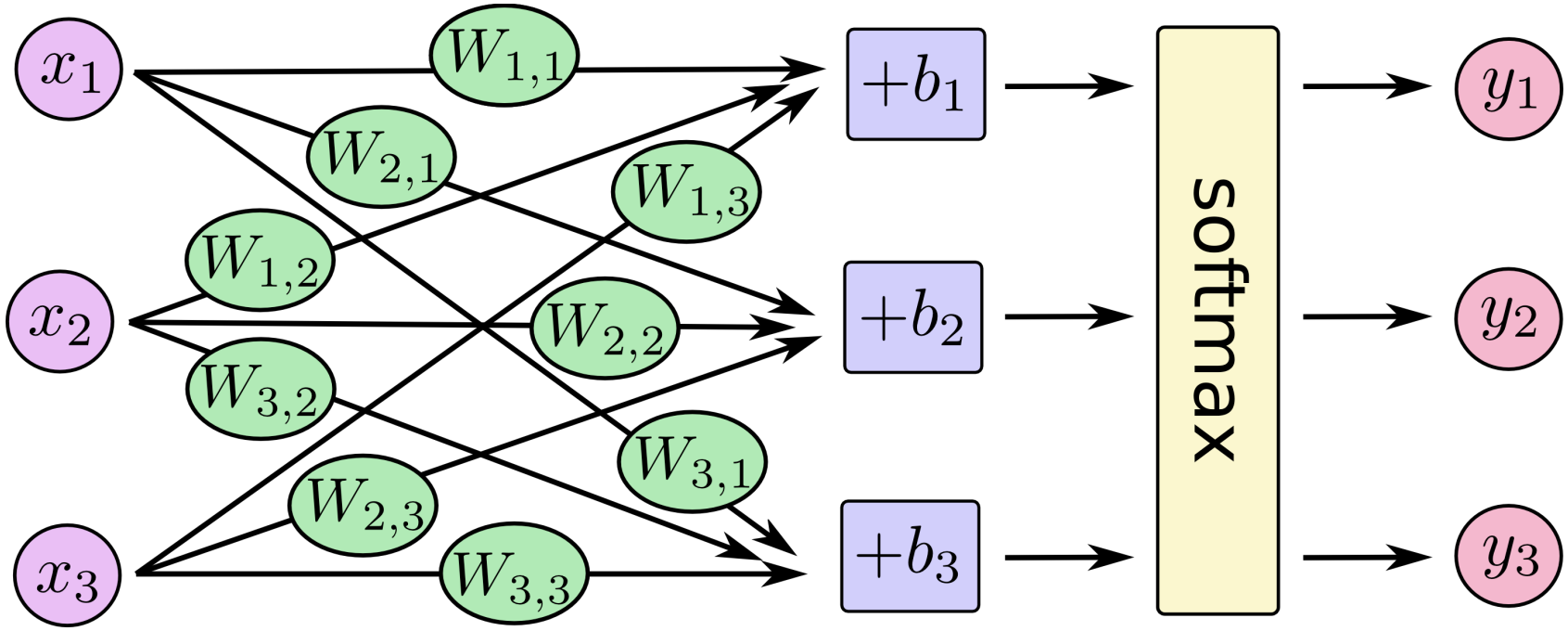
Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer NN	

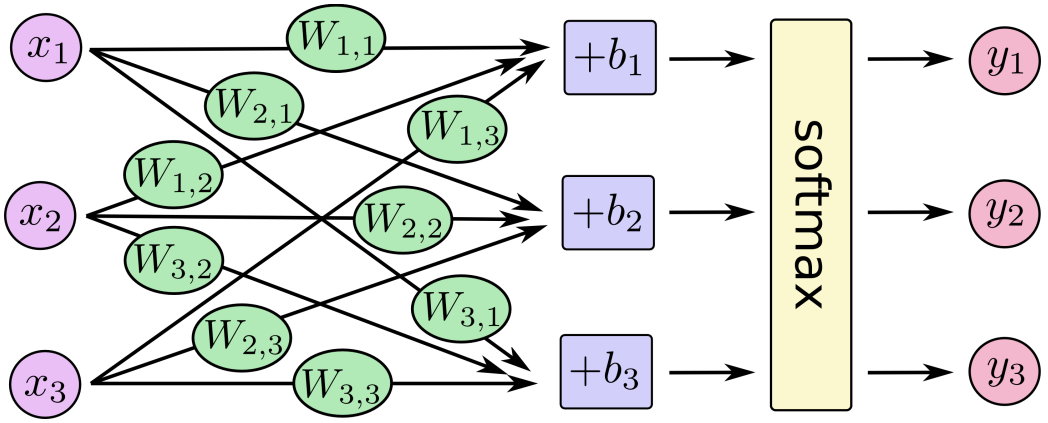
http://www.kdnuggets.com/2016/08/role-activation-function-neural-network.html



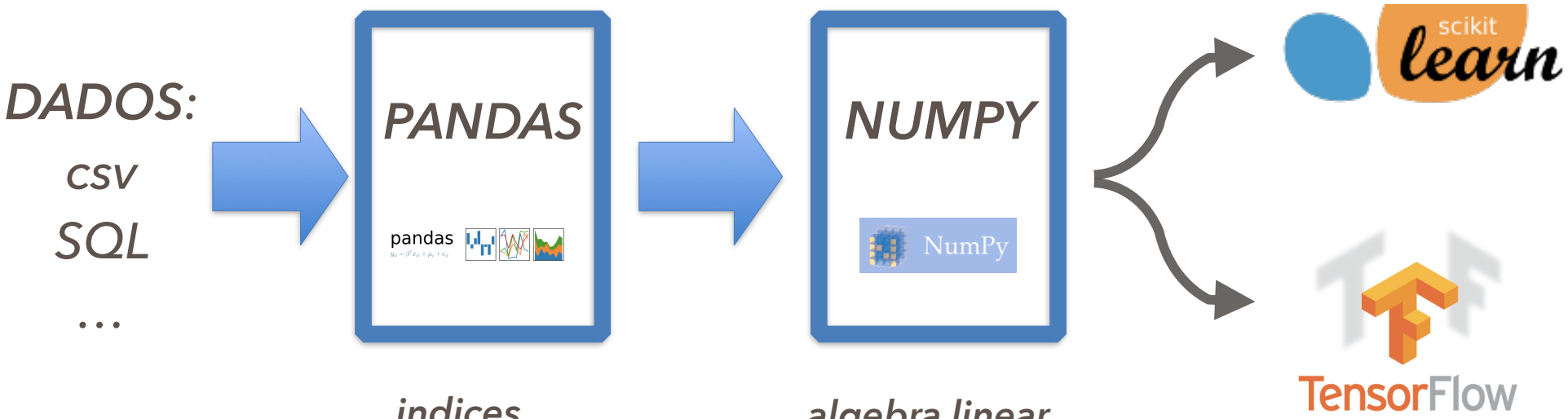
Este setup também é conhecido por
MULTINOMIAL LOGISTIC CLASSIFICATION

o que precisamos definir?





- arquitetura:
  - dimensão da entrada
  - dimensão da saída (numero de classes)
  - dimensões escondidas - # de pesos & bias
  - função de ativação
- operação:
  - tamanho do batch
  - qtd de epochs
- objetivos:
  - função de custo & métrica
  - otimizador



indices
slicing, indexing
limpeza
get_dummies
merge, concat
groupby

algebra linear
multiplicacao matrizes
prod. interno
transform. linear

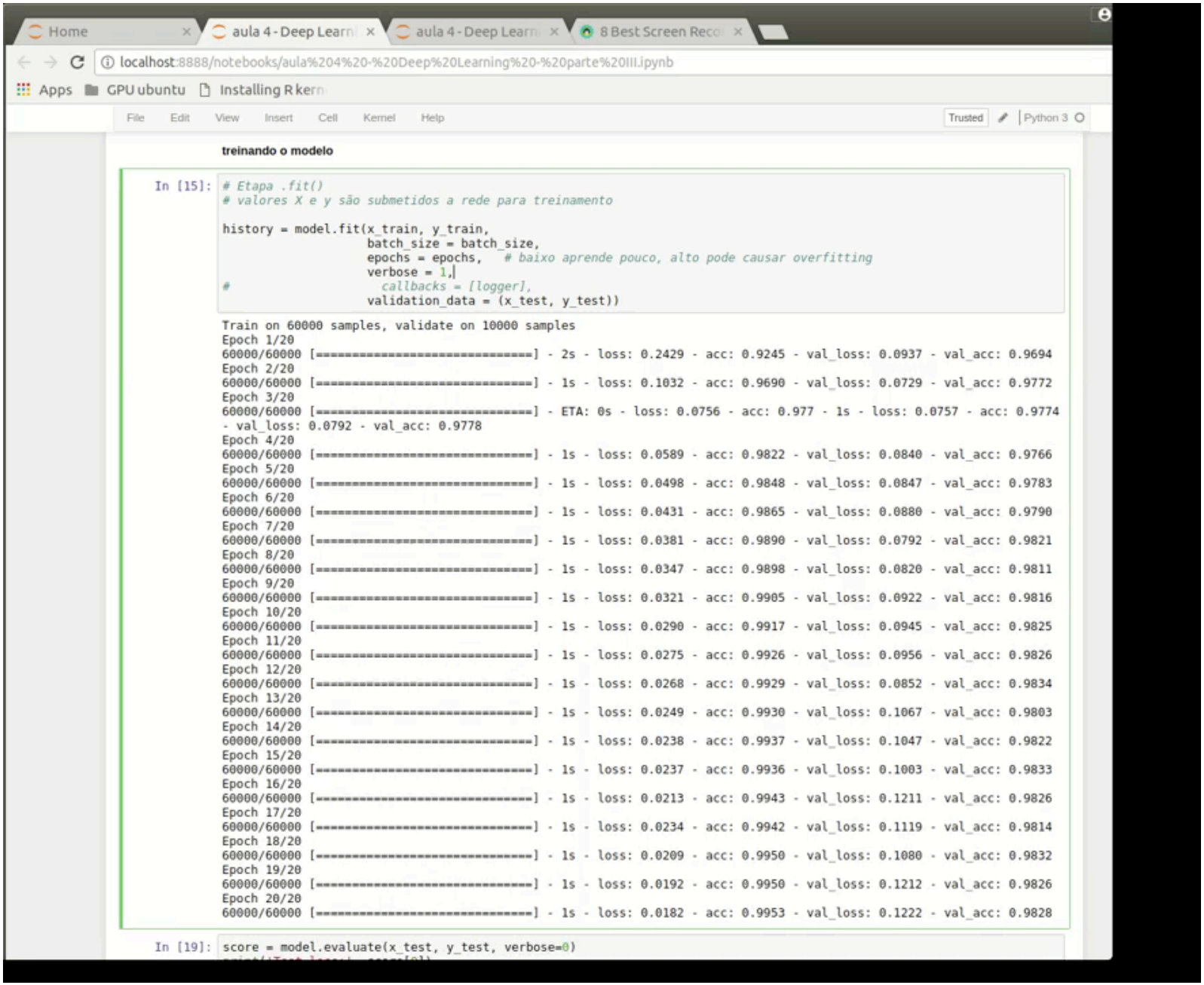
SGD 'BLACK MAGIC'

MANY HYPER-PARAMETERS

- INITIAL LEARNING RATE
- LEARNING RATE DECAY
- MOMENTUM
- BATCH SIZE
- WEIGHT INITIALIZATION

CPU vs. GPU

FGV MANAGEMENT



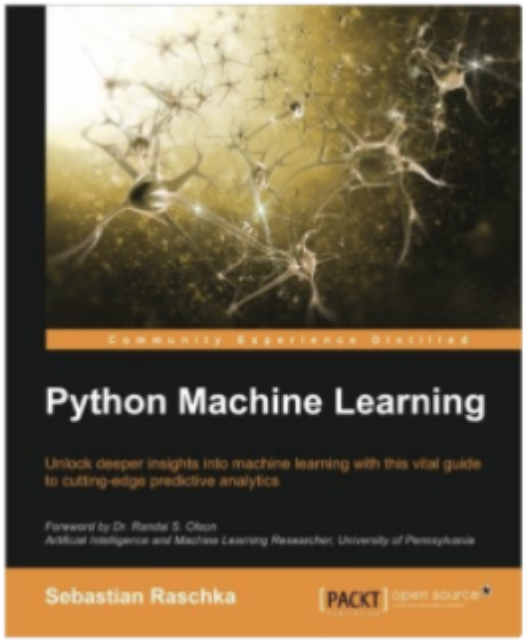
CPU vs. GPU

FGV MANAGEMENT

SISTEMAS DE RECOMENDAÇÃO

FGV MANAGEMENT

Look inside



Flip to back

See all 3 images

Python Machine Learning Paperback – September 23, 2015

by Sebastian Raschka (Author)

4.4 out of 5 stars 101 customer reviews

See all 2 formats and editions

Kindle	Paperback
$31.56	$40.49 Prime
Read with Our Free App	38 Used from $28.75 24 New from $35.00

Unlock deeper insights into Machine Leaning with this vital guide to cutting-edge predictive analytics

About This Book

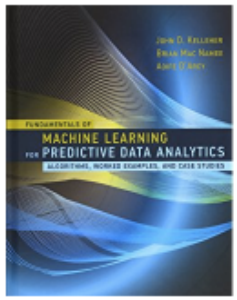
- Leverage Python's most powerful open-source libraries for deep learning, data wrangling, and data visualization
- Learn effective strategies and best practices to improve and optimize machine learning systems and algorithms
- Ask – and answer – tough questions of your data with robust statistical

Read more

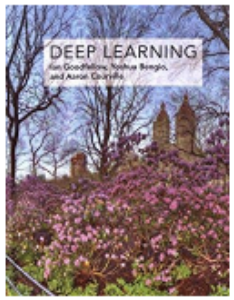
Report incorrect product information.

Customers who bought this item also bought

Page 1 of 25



Fundamentals of Machine Learning for Predictive Data Analytics:...
› John D. Kelleher
4.4 out of 5 stars 23
Hardcover
$70.00 Prime



Deep Learning (Adaptive Computation and Machine Learning series)
› Ian Goodfellow
4.8 out of 5 stars 63
Hardcover
$72.00 Prime

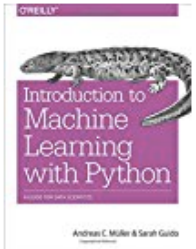
Data Science from Scratch: First Principles with Python
› Joel Grus
4.1 out of 5 stars 80
#1 Best Seller in Data Mining
Paperback
$28.97 Prime

Hands-On Machine Learning with Scikit-Learn and TensorFlow:...
› Aurélien Géron
5.0 out of 5 stars 3
#1 Best Seller in Computer Vision & Pattern...
Paperback
$30.10 Prime

Your recently viewed items and featured recommendations

Inspired by your browsing history

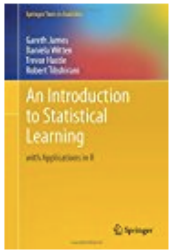
Page 1 of 10



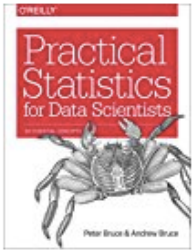
Introduction to Machine Learning with Python:...
Andreas C. Müller
4.2 out of 5 stars 14
Paperback
$34.38 Prime

Data Science from Scratch: First...
› Joel Grus
4.1 out of 5 stars 80
Paperback
$28.97 Prime

Python Data Science Handbook: Essential...
› Jake VanderPlas
4.1 out of 5 stars 12
Paperback
$45.47 Prime



An Introduction to Statistical Learning:...
› Gareth James
4.5 out of 5 stars 139
Hardcover
$70.35 Prime



Practical Statistics for Data Scientists: 50 Essential Concepts
› Peter Bruce
Paperback
$22.85 Prime

SISTEMAS DE RECOMENDAÇÃO

FGV MANAGEMENT

- técnicas, ferramentas e sistemas…
- … para sugerir ítens a pessoas

SISTEMAS DE RECOMENDAÇÃO
objetivos:

FGV MANAGEMENT

- aumentar vendas - volume
- aumentar vendas - diversidade
- aumentar satisfação
- aumentar fidelização
- entender melhor o que o cliente quer

SISTEMAS DE RECOMENDAÇÃO
por que?

FGV MANAGEMENT

- **Cauda Longa:** "We are leaving the age of information and entering the age of recommendation" - Chris Anderson: "The Long Tail…"
- **Big Data:** "People read 10MB a day, hear 400MB a day and see 1MB every second" - The economist Nov/2006
- **Paradoxo da Escolha:** "I want a pair of jeans"; "Do you want stone-washed, acid-washed, or distressed? Do you want button-fly or zipper-fly? Regular or faded? - Barry Schwartz : "Paradox of choice, why more is less"

SISTEMAS DE RECOMENDAÇÃO
cases

FGV MANAGEMENT

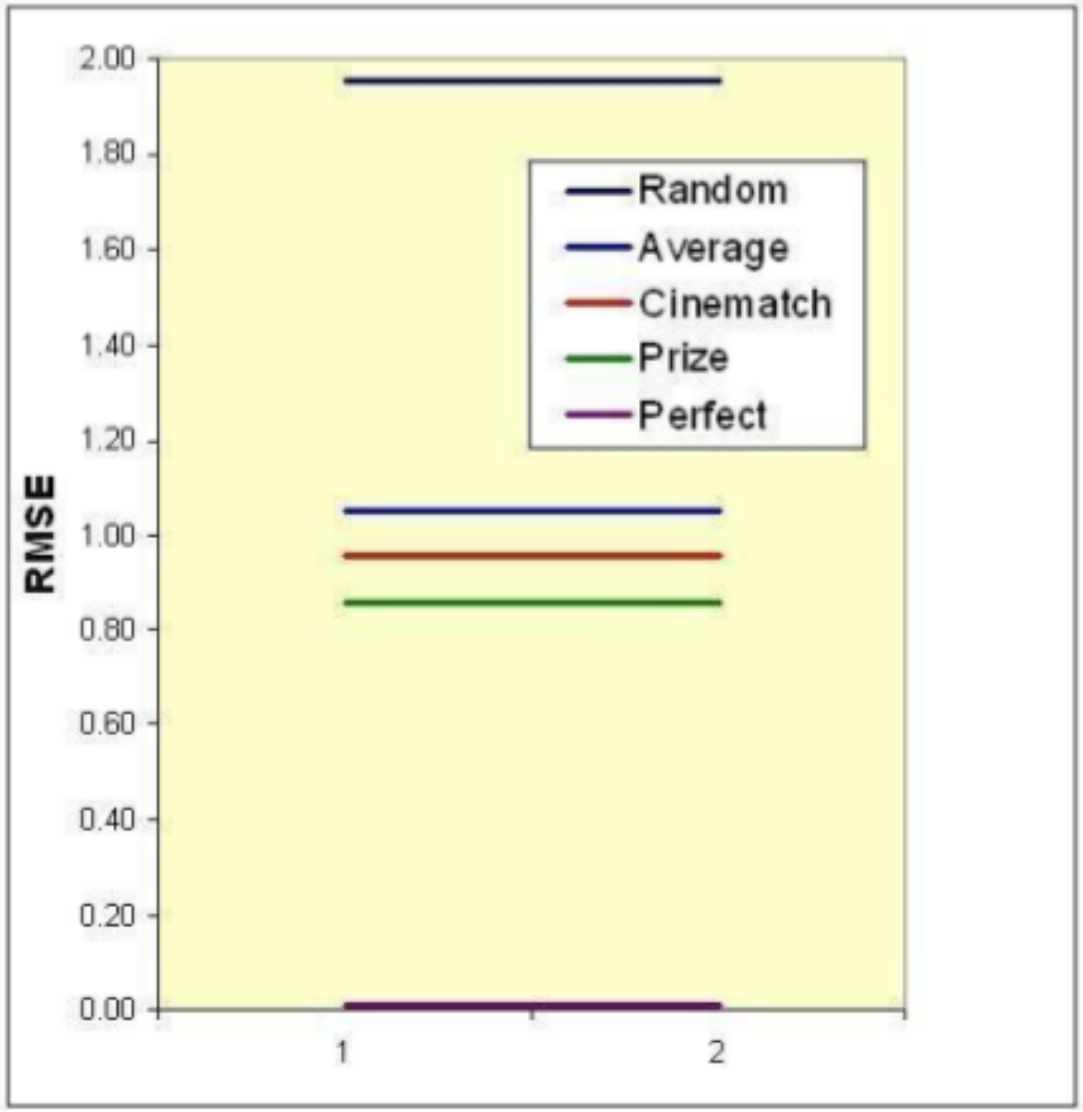
- **Netflix:** mais de 60% dos filmes assistidos foram recomendados
- **Google news**: SR geram 38% mais *clickthrough*
- **Amazon**: 35% das vendas vem dos SR

- recomendação: item ➜ pessoa
- busca: pessoa ➜ item
- a recomendação “acha” a pessoa
- busca e recomendação: dois lados de uma mesma moeda

- Função que prediz se uma pessoa vai gostar de um item
- dados:
  - similaridade: outras pessoas & feedback passado
  - similaridade: outros itens & feedback passado
  - contexto
  - …

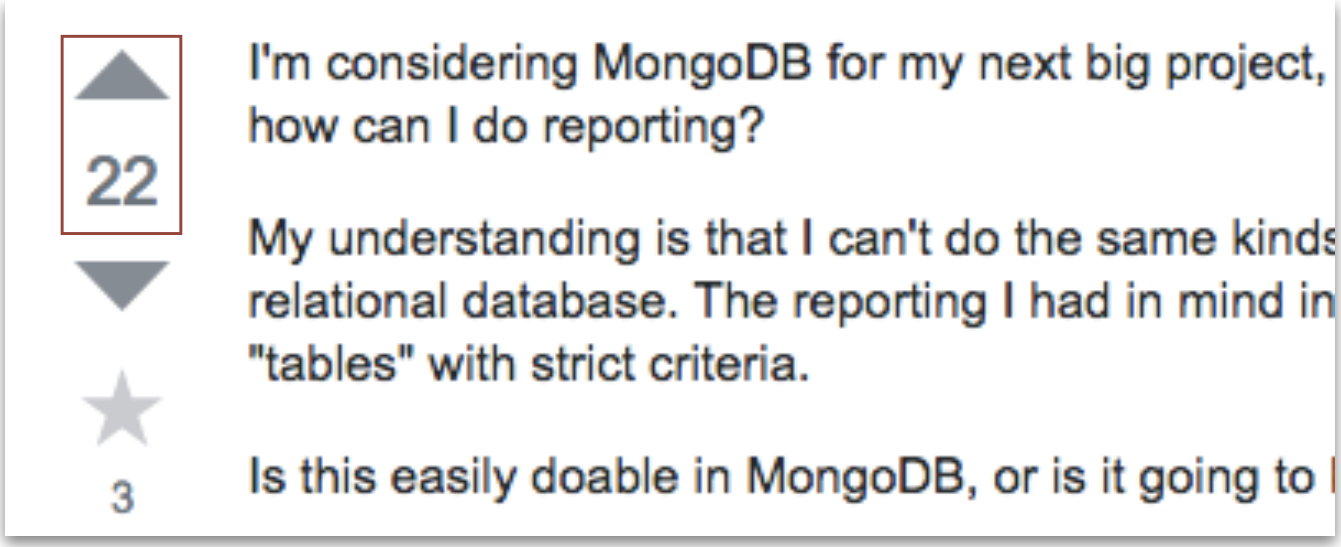
complexidade VS resultado

FGV MANAGEMENT



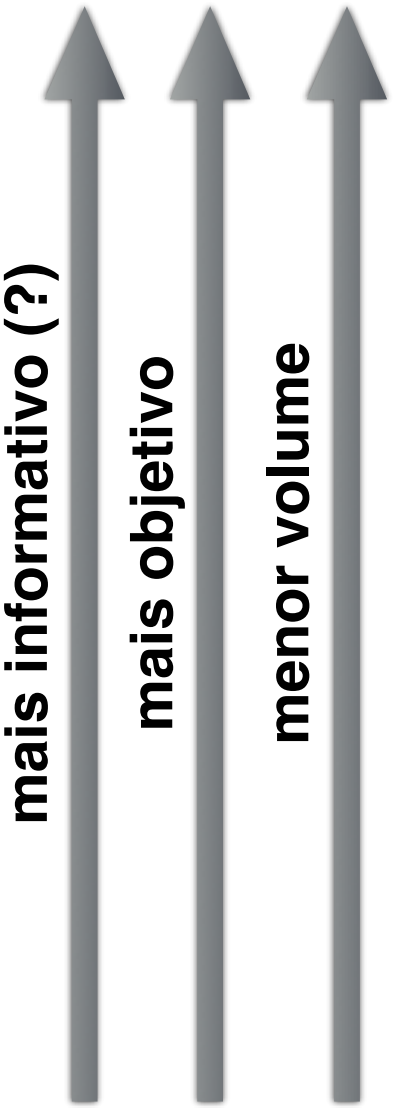
tipos de dados
(expressando preferências)

FGV MANAGEMENT

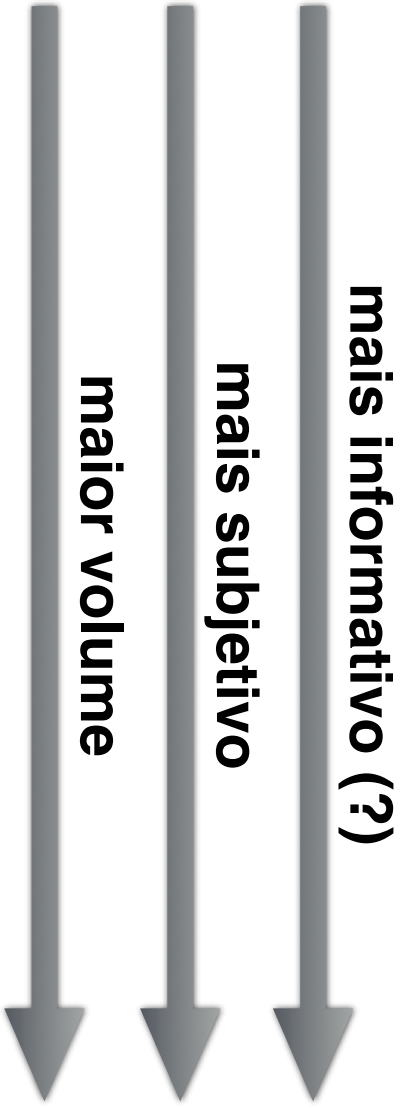


tipos de dados
(expressando preferências)

FGV MANAGEMENT



- Explícitos:
  - rating
  - avaliação
  - voto
- Implícitos:
  - click
  - compra
  - follow



implícito versus explicito

FGV MANAGEMENT

- qual é melhor ?
- para recomendações
- explicito: avaliações

  implícito: intenções
- gostar != recomendar

Clip slide

Approaches to Recommendation

- non-personalized
- Collaborative Filtering: Recommend items based only on the users past behavior
  - **User-based:** Find similar users to me and recommend what they liked
  - **Item-based:** Find similar items to those that I have previously liked
- Content-based: Recommend based on item features

Fonte: https://www.slideshare.net/xamat/recommender-systems-machine-learning-summer-school-2014-cmu

Clip slide

What works

- Depends on the **domain** and particular **problem**
- However, in the general case it has been demonstrated that the best isolated approach is CF.
  - Other approaches can be hybridized to improve results in specific cases (cold-start problem...)
- What matters:
  - Data preprocessing: outlier removal, denoising, removal of global effects (e.g. individual user's average)
  - "Smart" dimensionality reduction using MF/SVD
  - Combining methods

NETFLIX

Xavier Amatriain – July 2014 – Recommender Systems

Fonte: https://www.slideshare.net/xamat/recommender-systems-machine-learning-summer-school-2014-cmu

Fonte: https://www.slideshare.net/xamat/recommender-systems-machine-learning-summer-school-2014-cmu

relembrando

Algebra Linear

FGV MANAGEMENT

- para recomendar, podemos encontrar:
  - alguém similar
  - produto similar
- … e fazer a recomendação
- como medimos similaridade?

similaridade

FGV MANAGEMENT

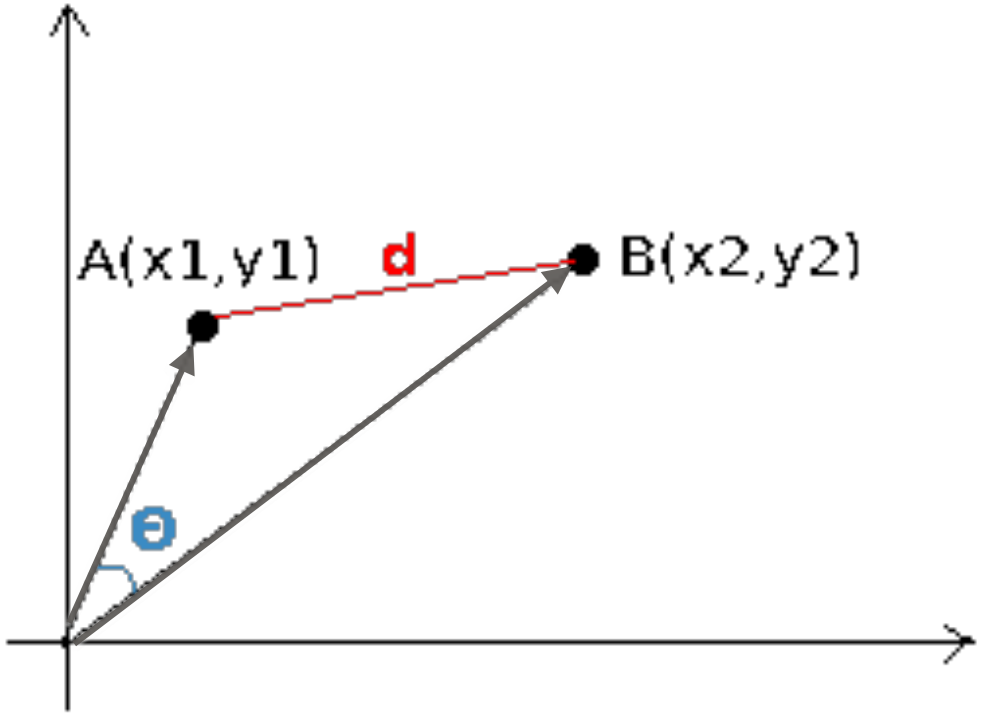
- distancia euclideana
- cosine similarity (similaridade de cosseno)
- pearson's r

FGV MANAGEMENT

- vetor
- produto interno
- angulo e cosseno do angulo
- distancia euclideana
- norma L2

Distancia euclideana

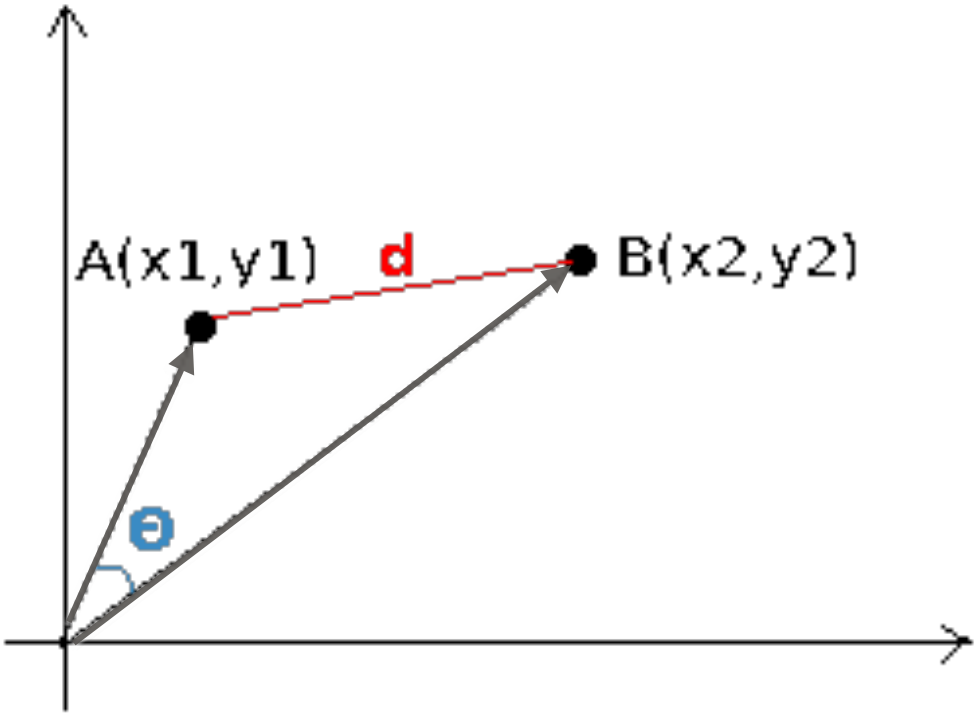
	x	y
A	1	3
B	4	4



algebra linear !
similaridade cosseno

FGV MANAGEMENT


	x	y
A	1	3
B	4	4



pearson's R



	x	y	z
A	1	3	1
B	4	4	6
C	2	2	3

DISTANCIA EUCLIDEANA

$$d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

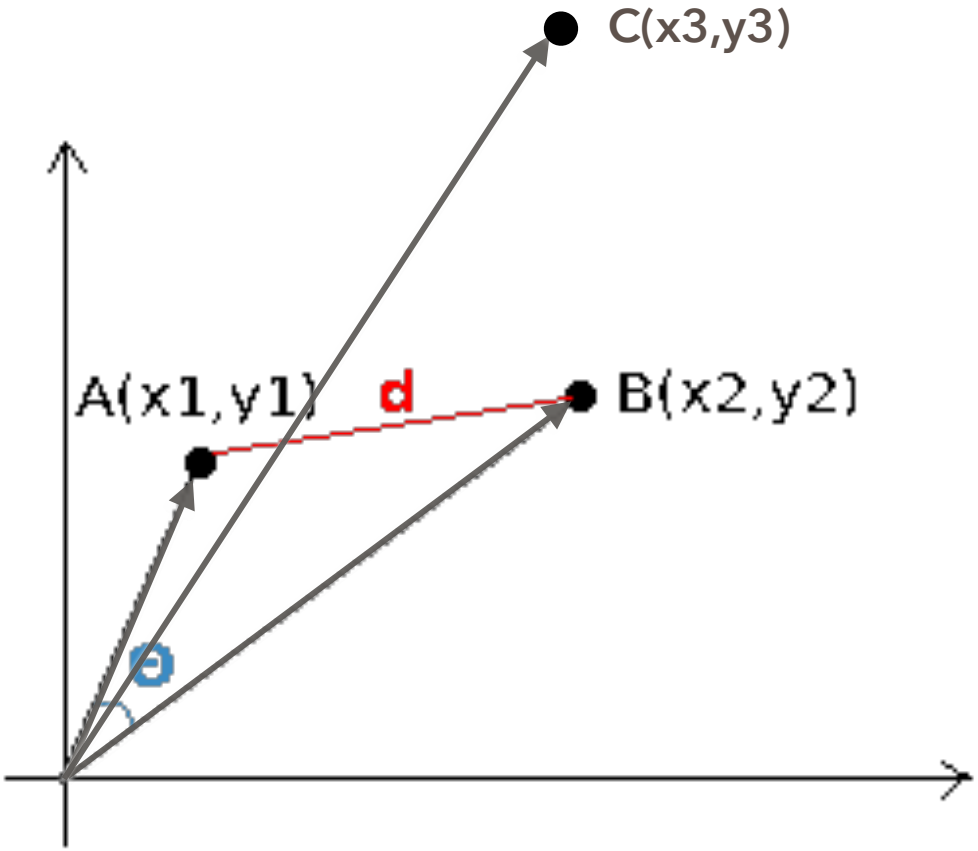
SIMILARIDADE COSSENO

$$cos(\theta) = \frac{\sum x_i * y_i}{\sqrt{\sum x_i^2} * \sqrt{\sum y_i^2}}$$

PEARSON'S R

$$\rho_{xy} = \frac{\sum((x_i - \bar{x}) * (y_i - \bar{y}))}{\sqrt{\sum(x_i - \bar{x})^2} * \sqrt{\sum(y_i - \bar{y})^2}}$$

Exercicio



		
A	1	3
B	4	4
C	4	7

Exercicio

Filtro Colaborativo

FGV MANAGEMENT

UB Collaborative Filtering

u_{ik}

- A collection of user u_i, $i=1, \ldots n$ and a collection of products p_j, $j=1, \ldots, m$
- An $n \times m$ matrix of ratings v_{ij}, with $v_{ij} = ?$ if user i did not rate product j
- Prediction for user i and product j is computed

$$v_{ij}^{*} = K \sum_{v_{kj} \neq ?} u_{jk} v_{kj} \quad \text{or} \quad v_{ij}^{*} = v_i + K \sum_{v_{kj} \neq ?} u_{jk} (v_{kj} - v_k)$$

- Similarity can be computed by Pearson correlation

$$u_{ik} = \frac{\sum_j (v_{ij} - v_i)(v_{kj} - v_k)}{\sqrt{\sum_j (v_{ij} - v_i)^2 \sum_j (v_{kj} - v_k)^2}} \quad \text{or} \quad \cos(u_i, u_j) = \frac{\sum_{k=1}^{m} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{m} v_{ik}^2 \sum_{k=1}^{m} v_{jk}^2}}$$

NETFLIX

Xavier Amatriain – July 2014 – Recommender Systems

https://grouplens.org/blog/similarity-functions-for-user-user-collaborative-filtering/

User-based CF Example

Clip slide

						sim(u,v)
	2		2	4	5	NA
	5		4			1
			5		2	
		1		5		4
			4			2
	4	5		1		NA

NETFLIX

User-based CF Example

						sim(u,v)
	2		2	4	5	NA
	5		4			1	0.87
			5		2		1
		1		5		4	-1
			4			2	
	4	5		1			NA

NETFLIX

Xavier Amatriain – July 2014 – Recommender Systems

User-based CF Example

						sim(u,v)
	2		2	4	5	NA
	5		4			1
			5		2	1
		1		5		4
	3.51*	3.81*	4	2.42*	2.48*	2
	4	5		1		NA

NETFLIX