

A aula interativa do Módulo 4 - Bootcamp Cientista de Dados começará em breve!

Atenção:

- 1) Você entrará na aula com o microfone e o vídeo DESABILITADOS.
- 2) Apenas a nossa equipe poderá habilitar seu microfone e seu vídeo em momentos de interatividade, indicados pelo professor.
- 3) Utilize o recurso Q&A para dúvidas técnicas. Nossos tutores e monitores estarão prontos para te responder e as perguntas não se perderão no chat.
- 4) Para garantir a pontuação da aula, no momento em que o professor sinalizar, você deverá ir até o ambiente de aprendizagem e responder a enquete de presença. Não é necessário encerrar a reunião do Zoom, apenas minimize a janela.

Técnicas para o Processamento do Big Data

Primeira Aula Interativa

Prof. Túlio Vieira

Técnicas para o Processamento do Big Data

Primeira Aula Interativa

Prof. Túlio Vieira

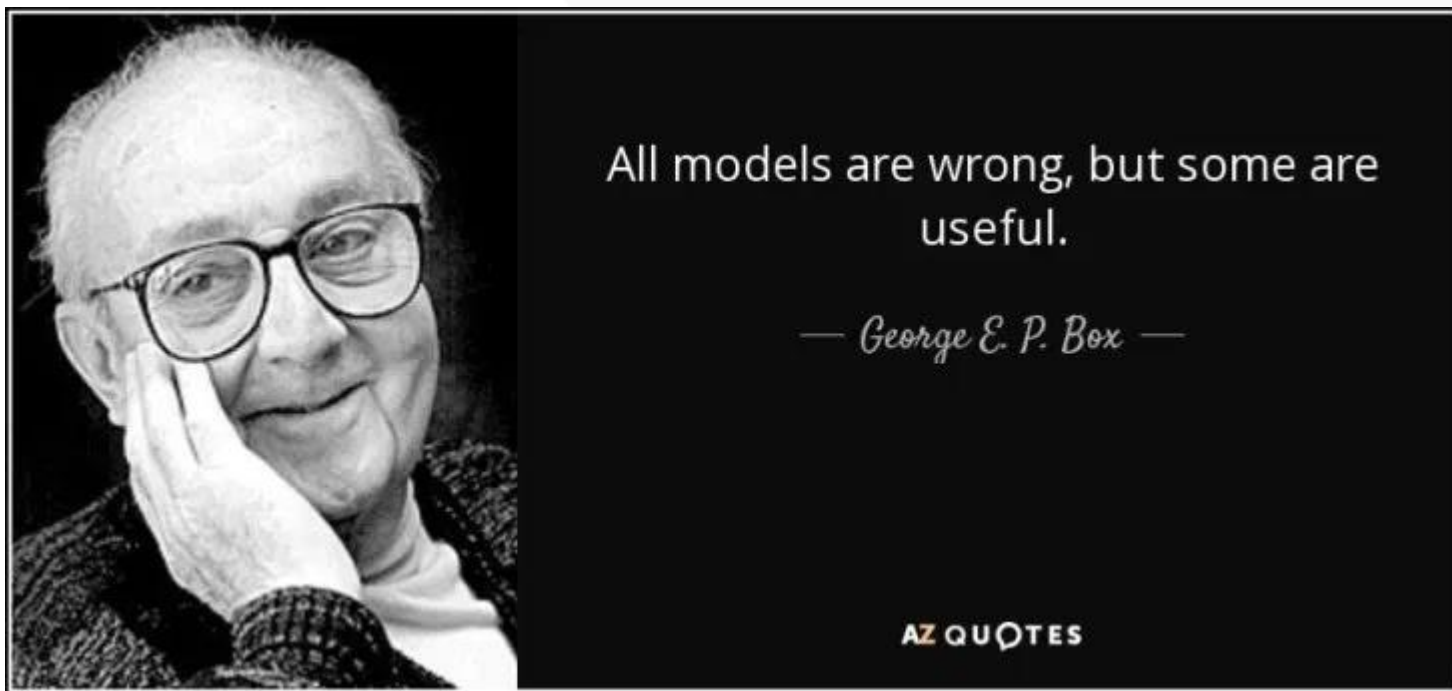


Nesta aula

1 Apresentação.

2 Correlação + PCA.

3 Prática.



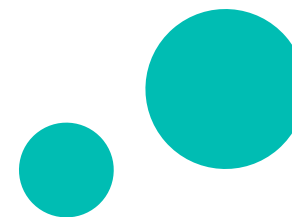


Apresentação

- Engenheiro de controle e automação pela UFMG, 2014
- Especialista em engenharia da qualidade pela PUC-MG, 2016;
- Mestre em Modelagem Matemática Computacional pelo CEFET-MG, 2017;
- Doutorando em Modelagem Matemática Computacional pelo CEFET-MG, previsão 2022.



Questão de Aquecimento

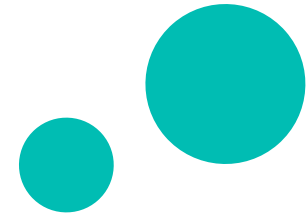


- Classificação Binária
 - Identificar se o empregado vai ou não continuar no emprego.
 - Dataset: Human Resources Analytics.

	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	promotion_last_5years	sales	salary	left
0	0.38	0.53	2	157	3	0	0	sales	low	1
1	0.80	0.86	5	262	6	0	0	sales	medium	1
2	0.11	0.88	7	272	4	0	0	sales	medium	1
3	0.72	0.87	5	223	5	0	0	sales	low	1
4	0.37	0.52	2	159	3	0	0	sales	low	1



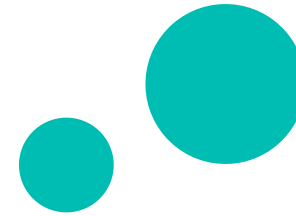
Questão de Aquecimento



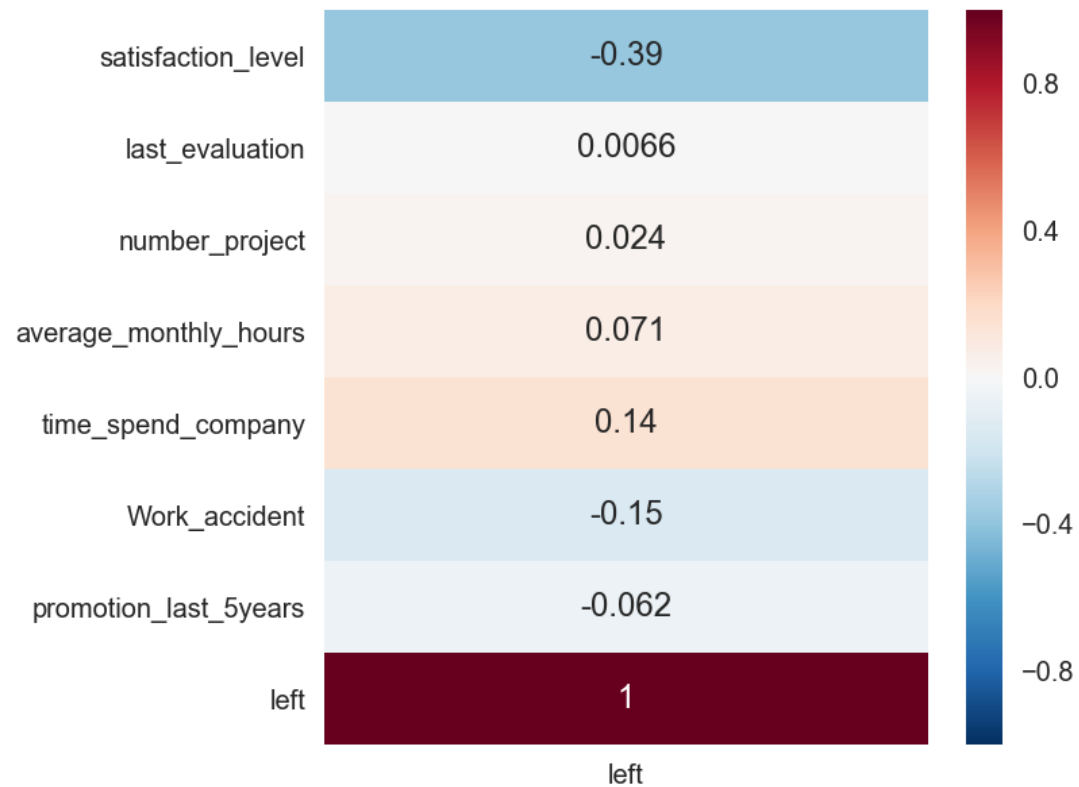
- Visualização e pré-processamento dos dados
 - Qual tipo de dados estamos trabalhando: valores reais, categóricos ou ambos?
 - Existem correlações entre as variáveis?
 - Existem dados nulos?
 - Existem dados que precisam ser normalizados?
 - Para os “labels”, existem classes não balanceadas ?

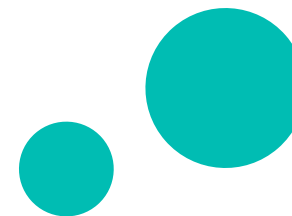


Questão de Aquecimento

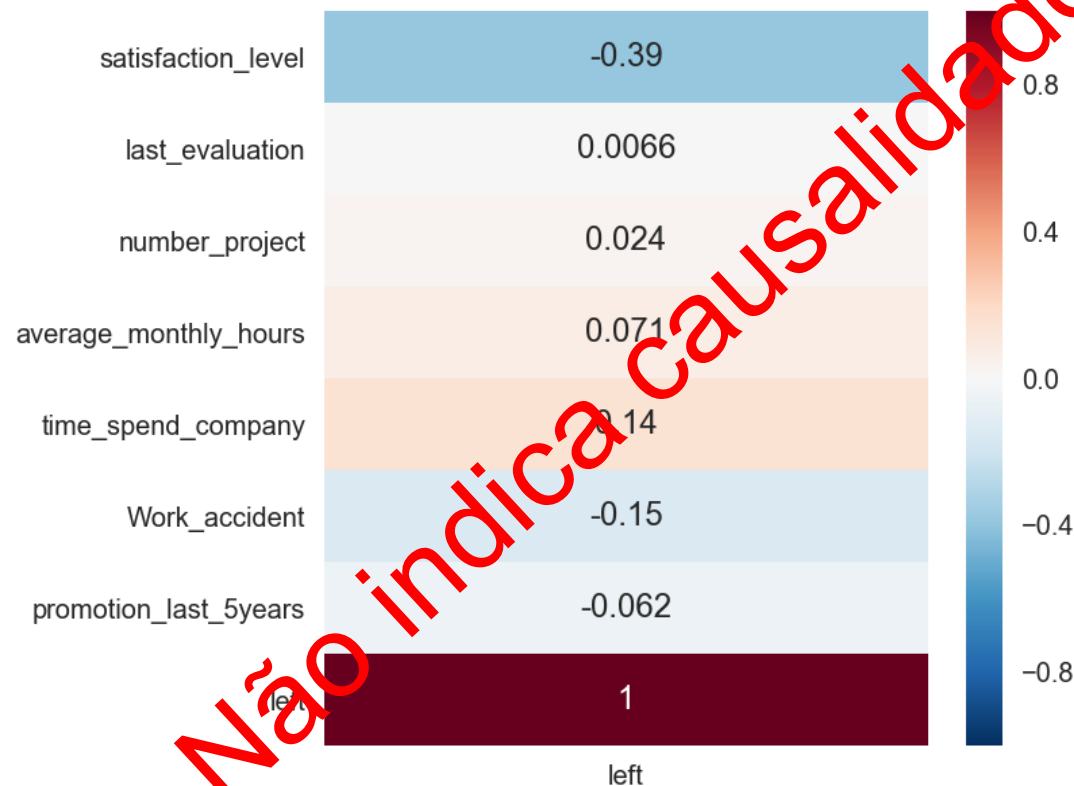


- Correlação





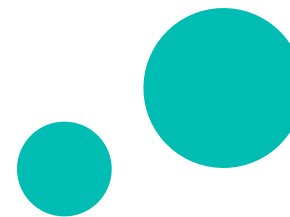
- Correlação



Não indica causalidade

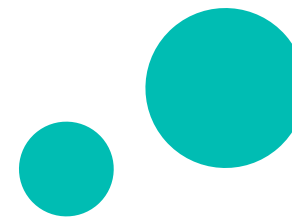


Questão de Aquecimento





Questão de Aquecimento

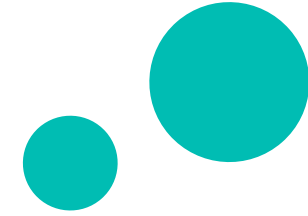


▪ Correlação



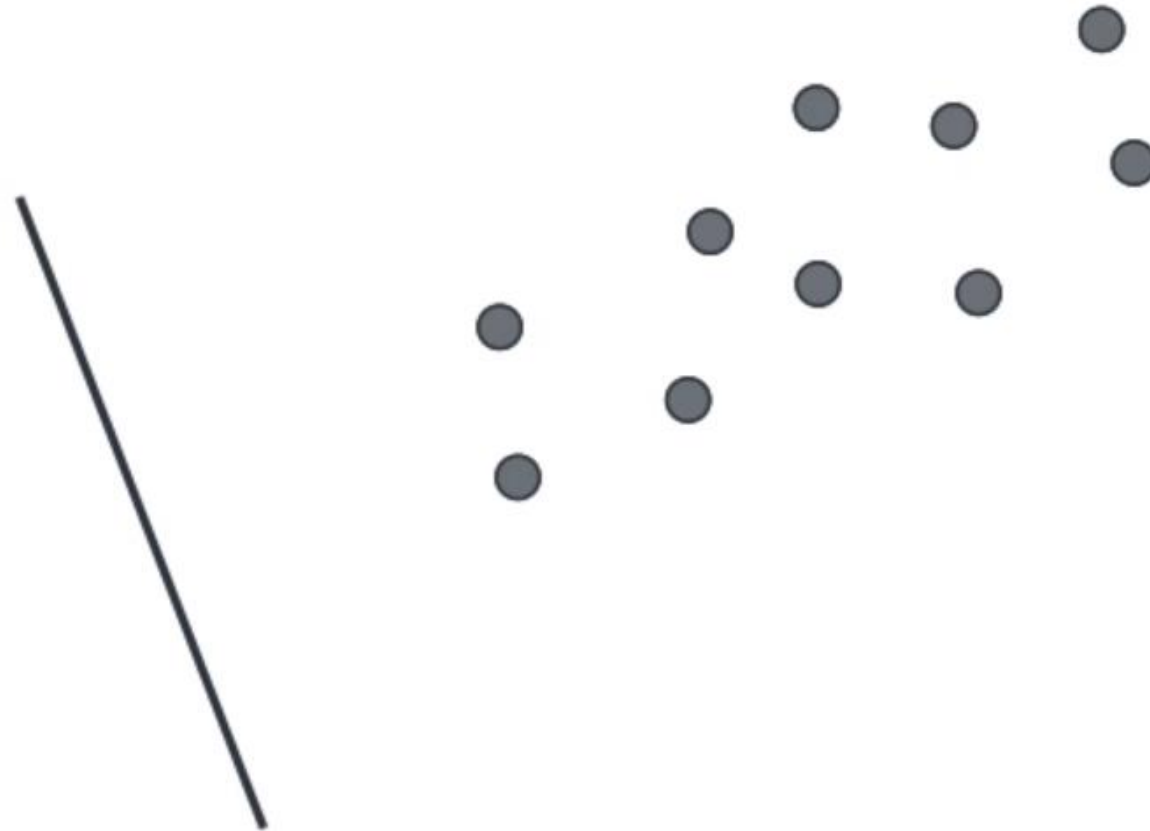
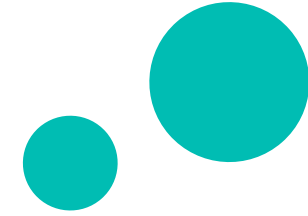


Principal Component Analysis (PCA)

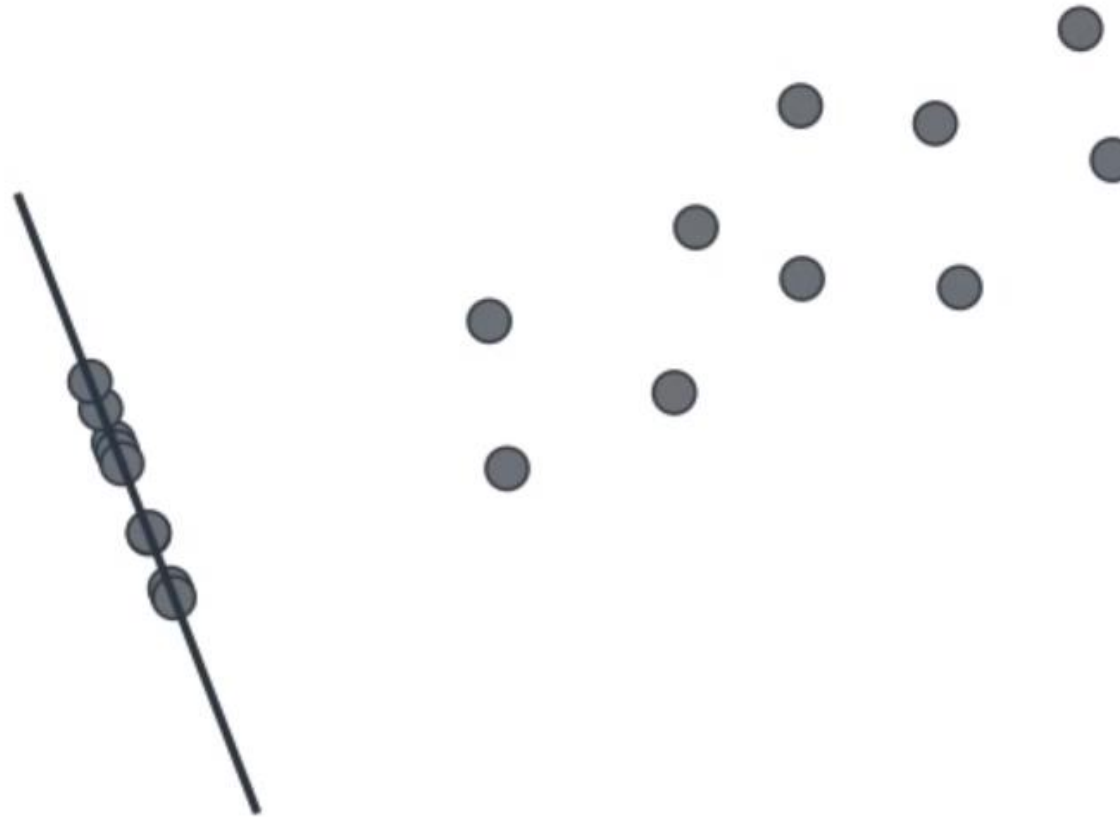
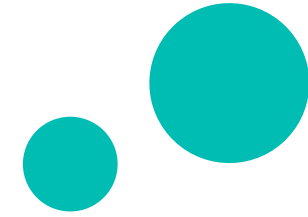


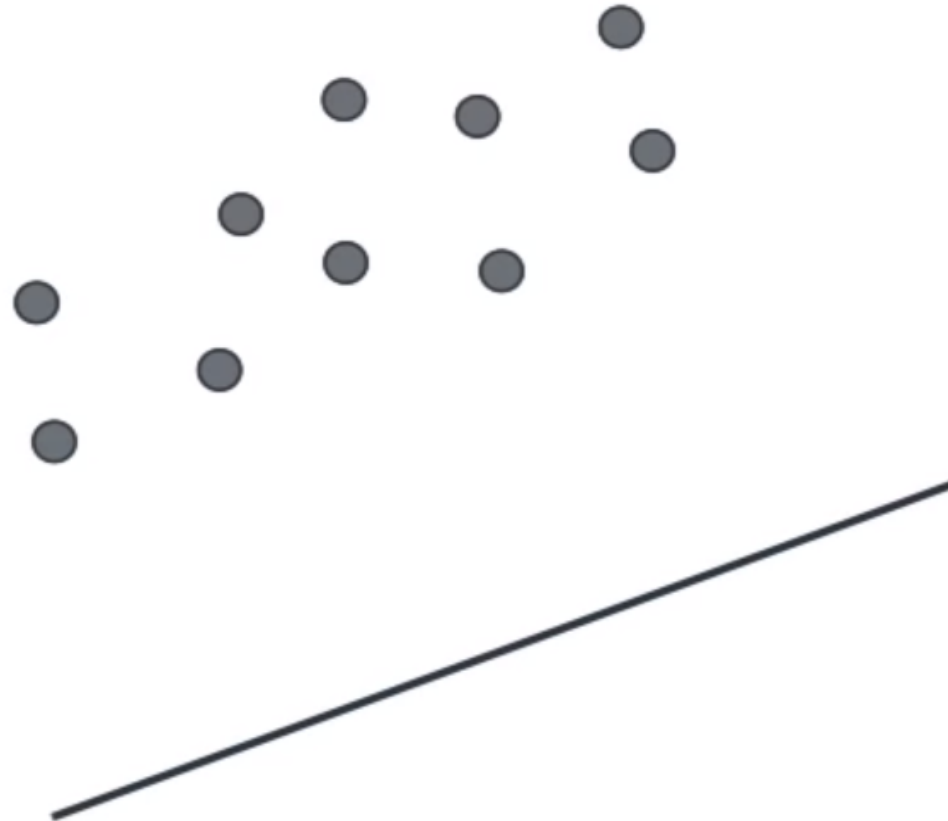
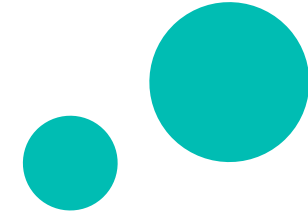


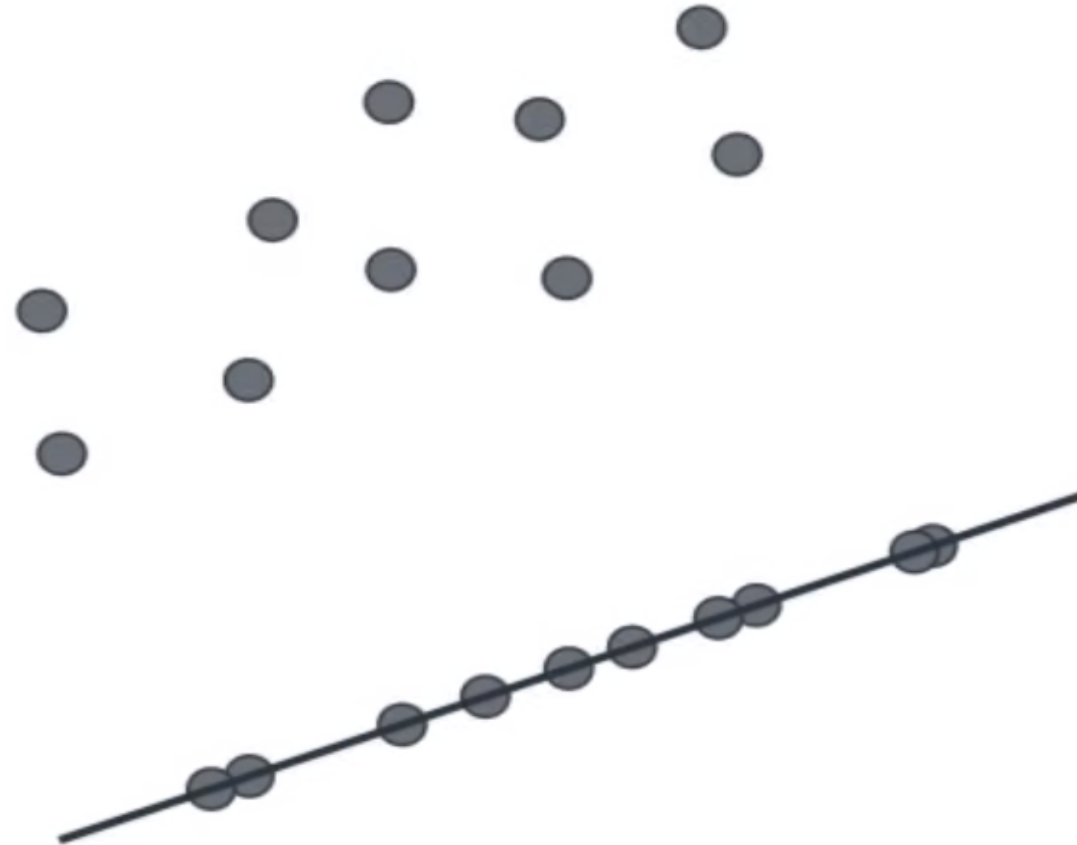
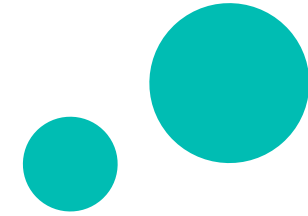
Principal Component Analysis (PCA)

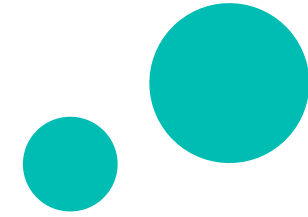


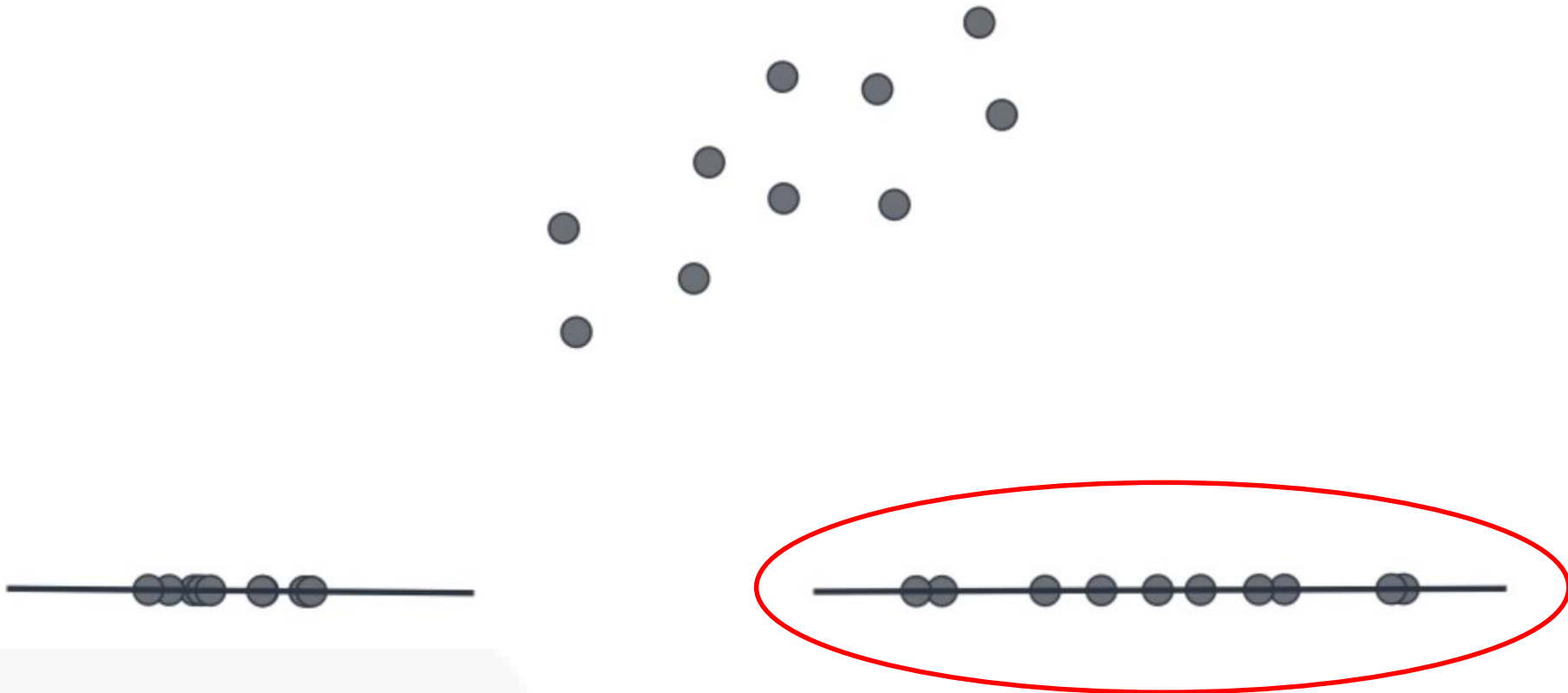
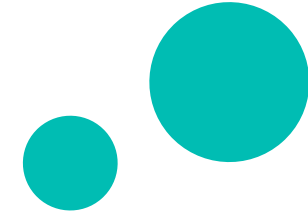
Principal Component Analysis (PCA)



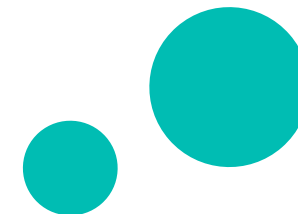






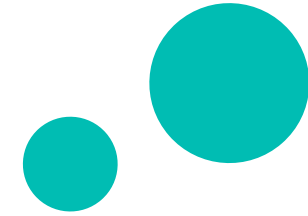


Principal Component Analysis (PCA)

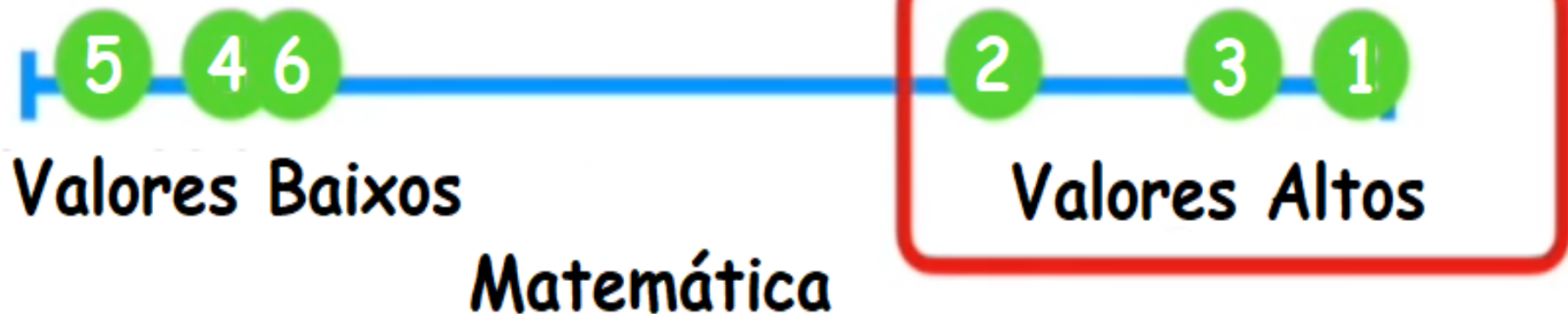


	Estudante 1	Estudante 2	Estudante 3	Estudante 4	...
Matemática	95	88	93	75	...
Português	96	79	98	81	...

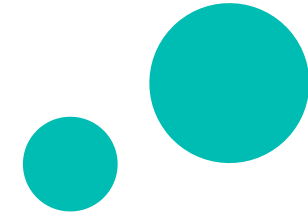
Principal Component Analysis (PCA)



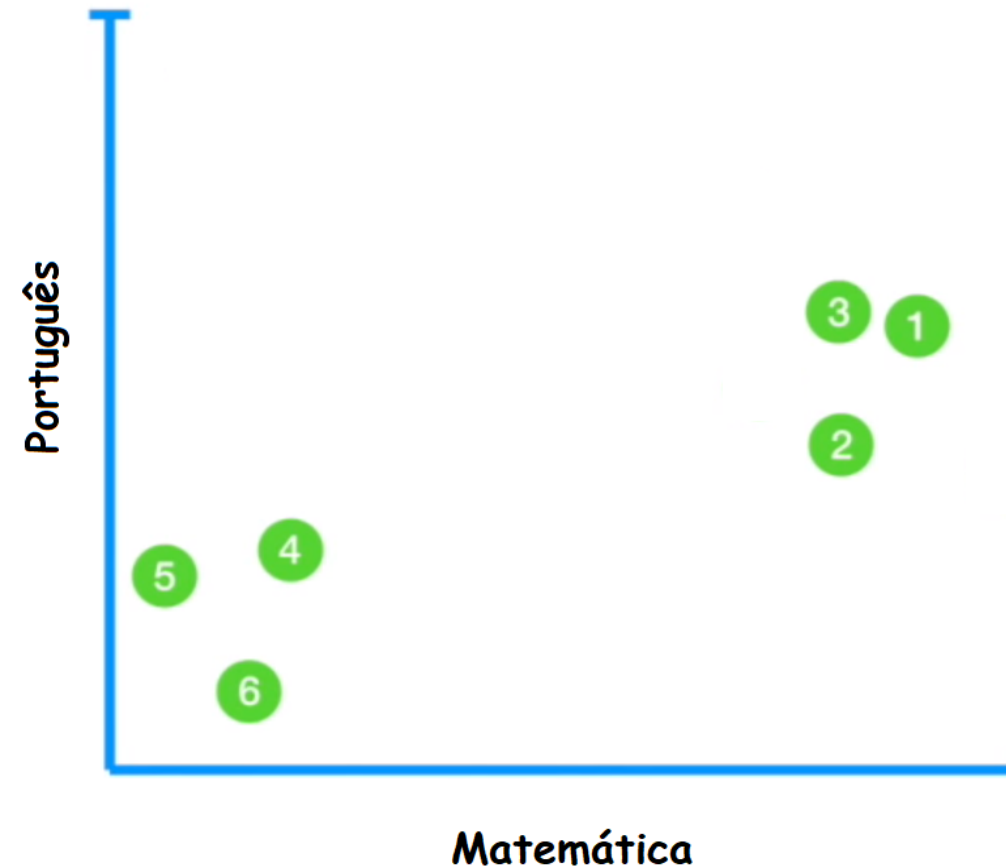
	Estudante 1	Estudante 2	Estudante 3	Estudante 4	...
Matemática	95	88	93	75	...
Português	96	79	98	81	...



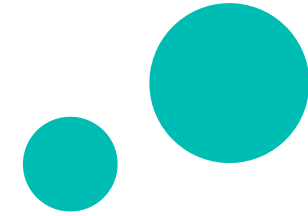
Principal Component Analysis (PCA)



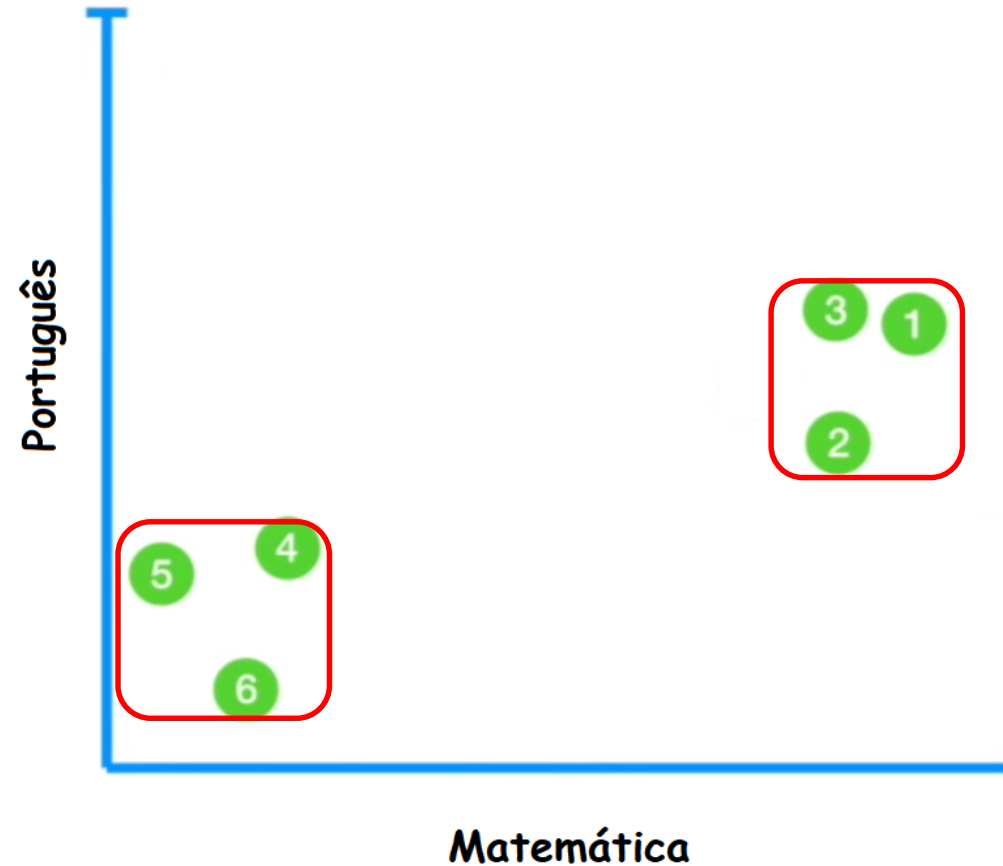
	Estudante	Estudante	Estudante	Estudante	...
	1	2	3	4	...
Matemática	95	88	93	75	...
Português	96	79	98	81	...



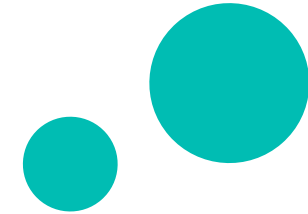
Principal Component Analysis (PCA)



	Estudante	Estudante	Estudante	Estudante	...
	1	2	3	4	...
Matemática	95	88	93	75	...
Português	96	79	98	81	...

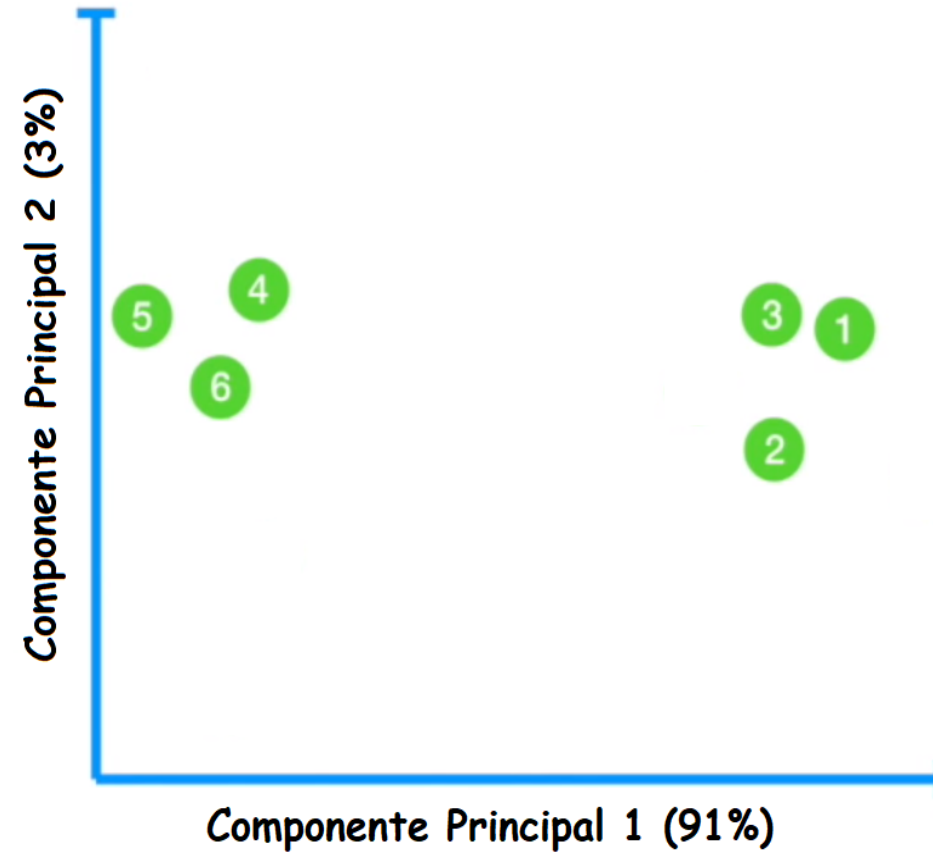


Principal Component Analysis (PCA)

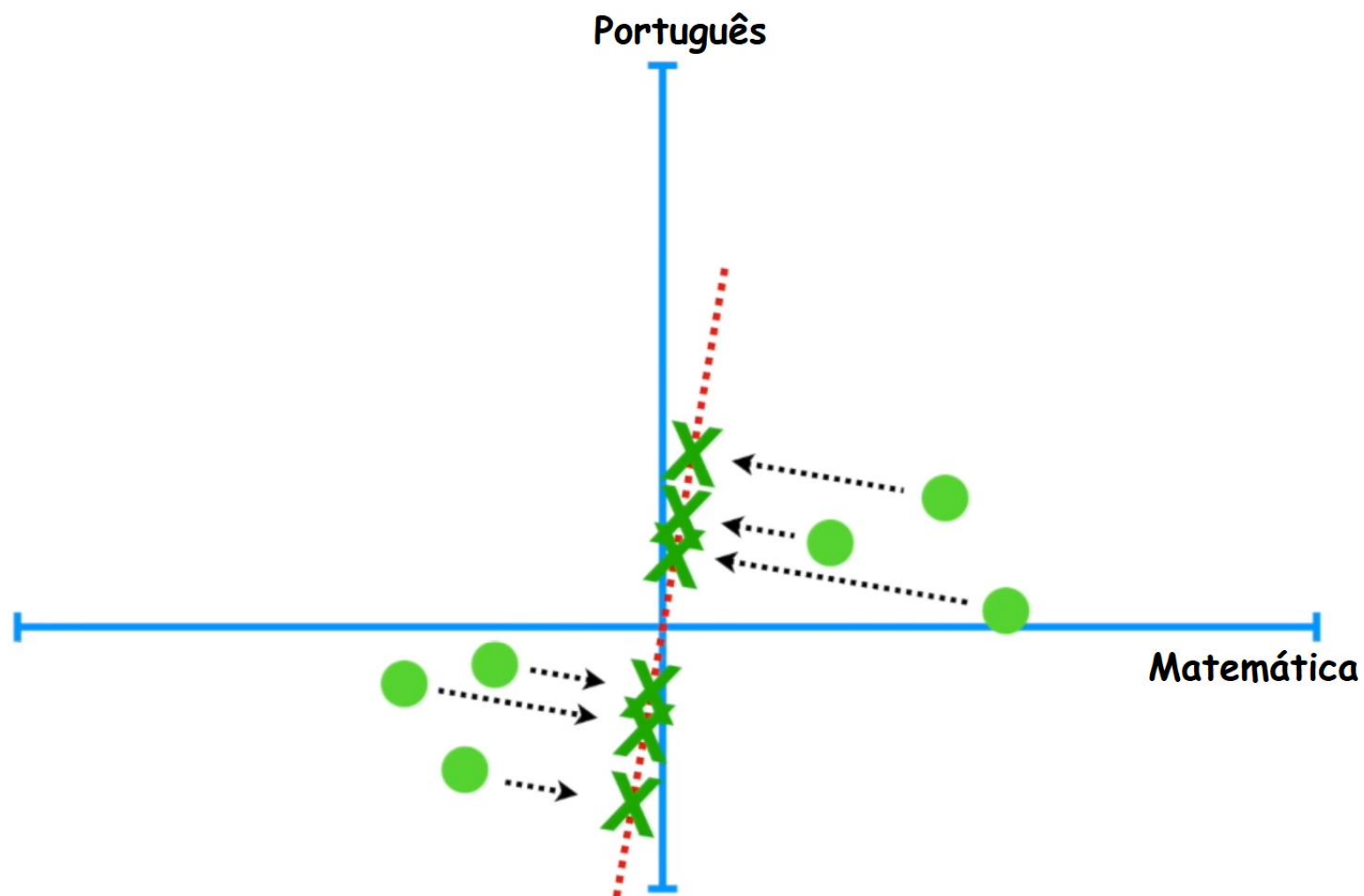


	Estudante	Estudante	Estudante	Estudante	...
	1	2	3	4	...
Matemática	95	88	93	75	...
Português	96	79	98	81	...

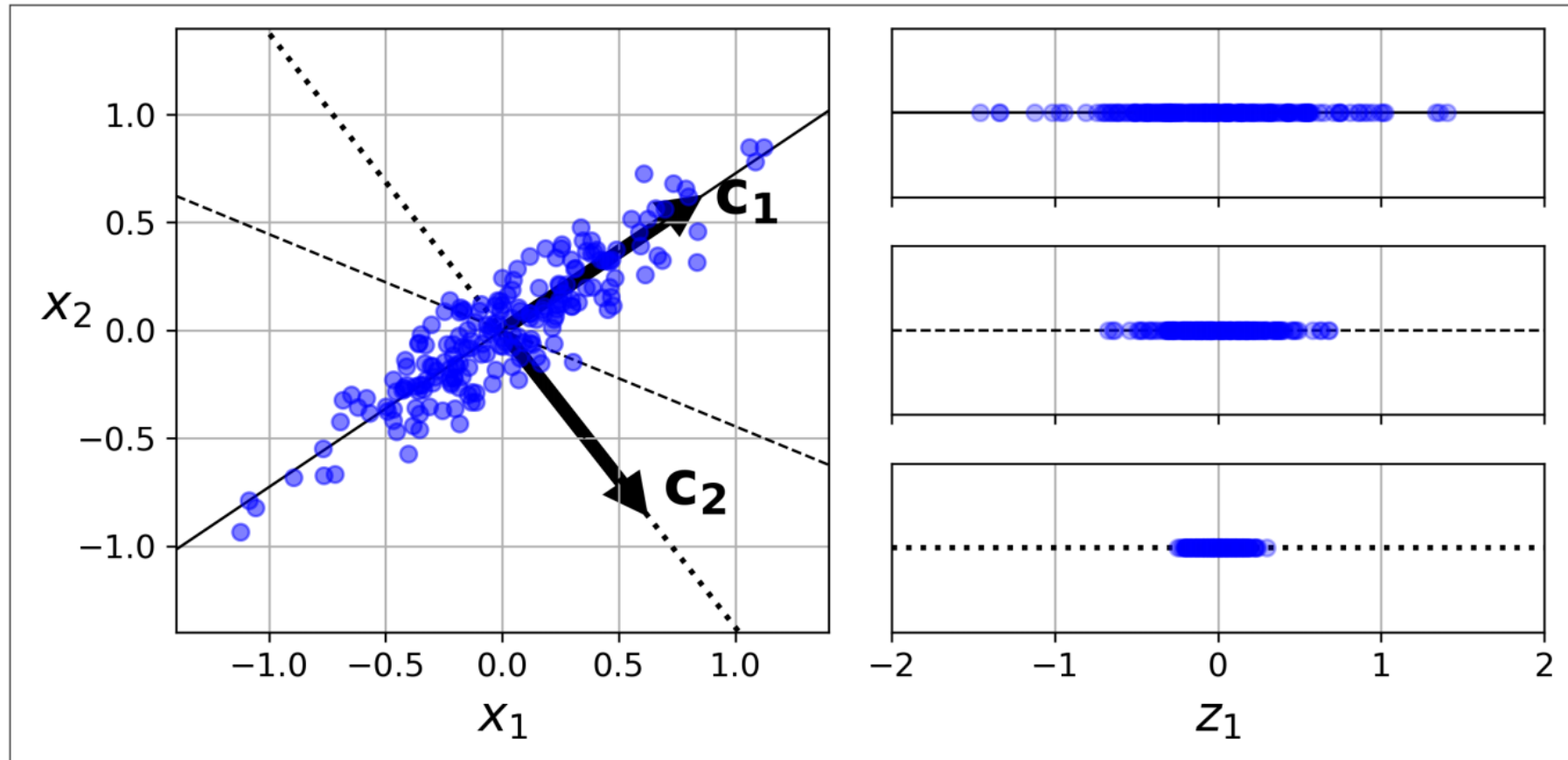
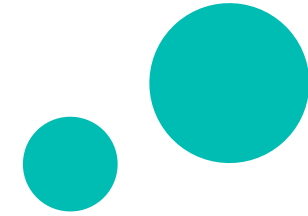
.

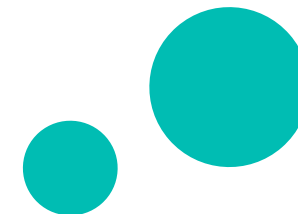


Principal Component Analysis (PCA)

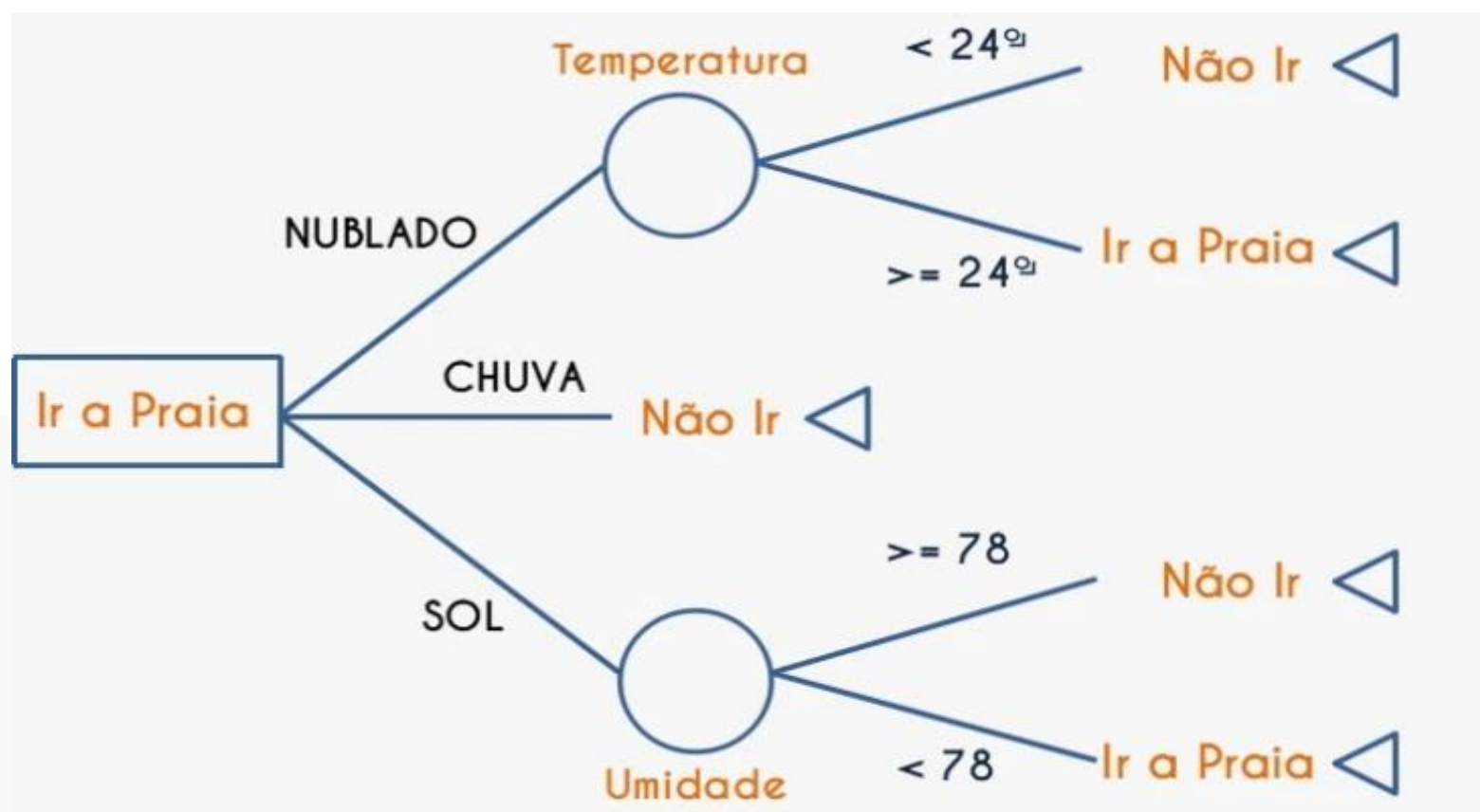


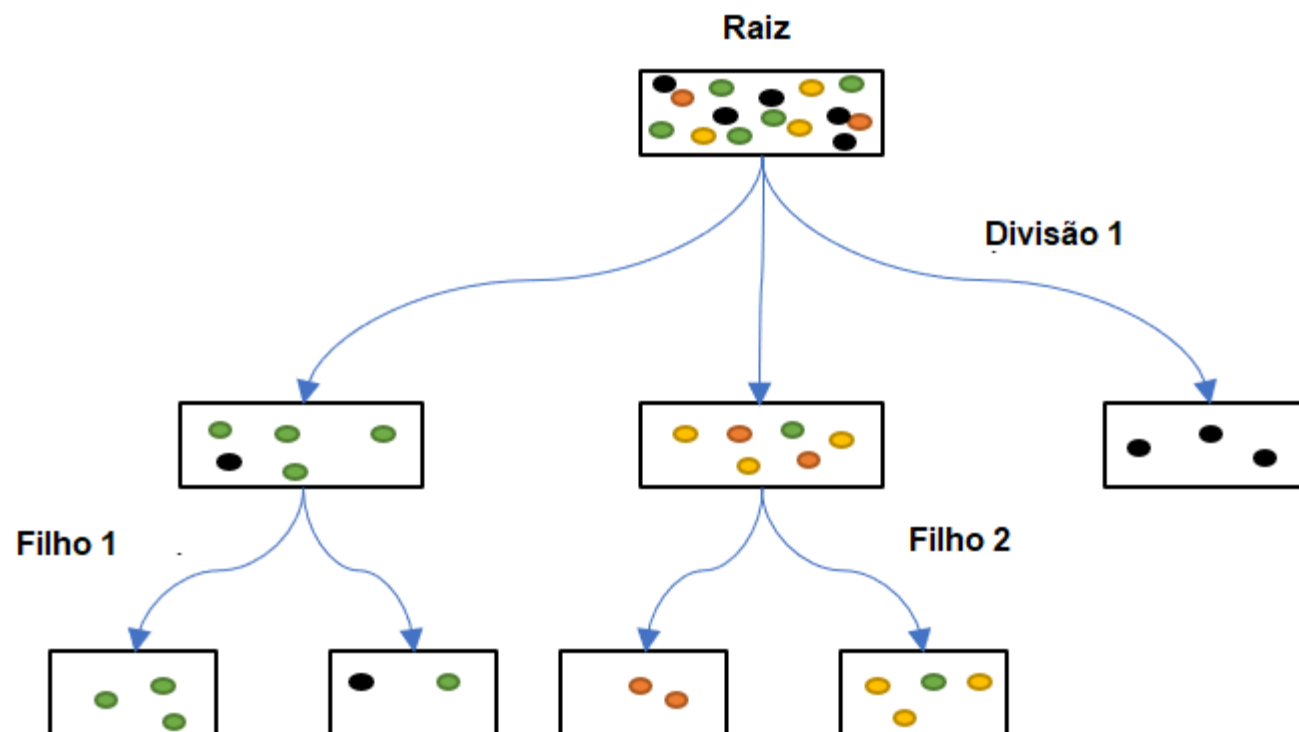
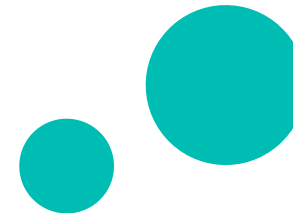
Principal Component Analysis (PCA)





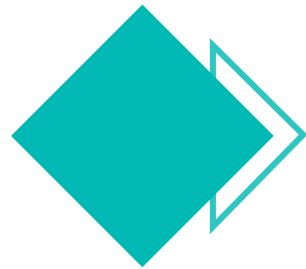
Vamos à Praia?



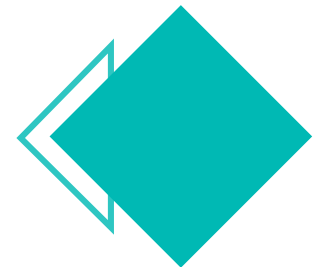
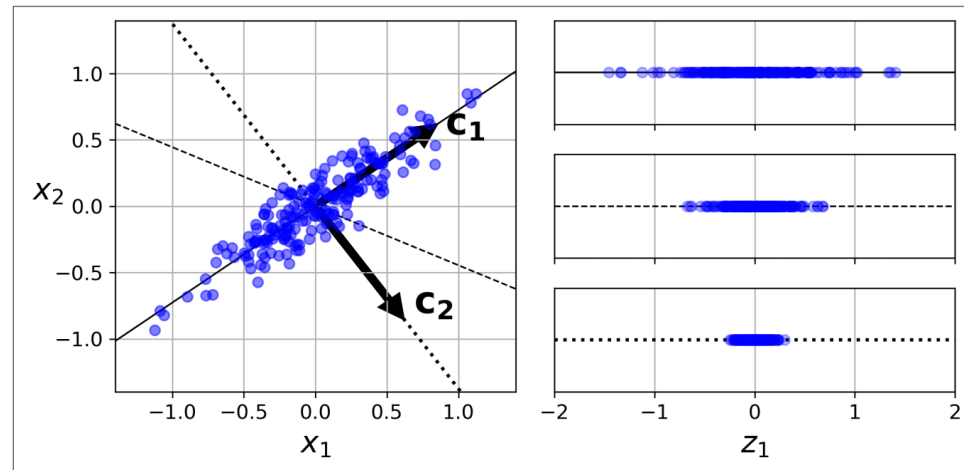


iGTi

Exemplo



colab





Na Aula de Hoje

1 Correlação+PCA.

2 Prática.



Próxima Aula

1 Dúvidas.

2 Desafio.