Máquinas de Vetor Suporte
(Support Vector Machines - SVM)

Lilianne Mariko Izuti Nakazono
Doutoranda em Astronomia (IAG-USP)

Support Vector Machines (SVM)

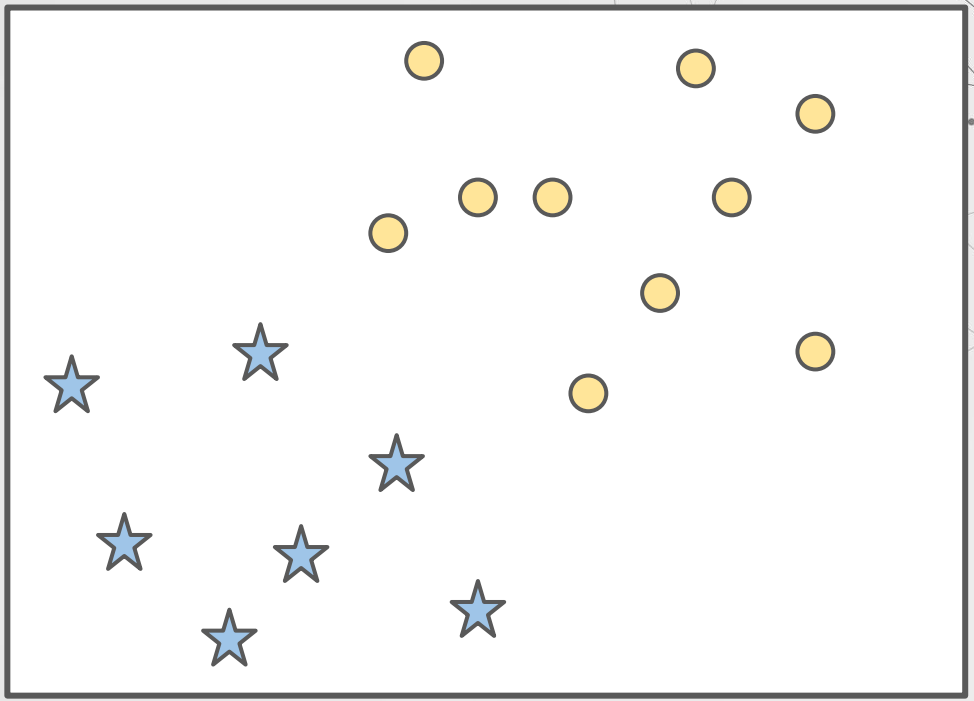
- Elaborado na década de 90 por V. Vapnik e colegas na AT&T Bell Laboratories
- O algoritmo encontra o melhor hiperplano que separa duas classes ao maximizar a distância entre os pontos mais próximos de cada classe (Vapnik, 1996)

Margem

Por simplicidade, vamos supor duas classes em um espaço de duas dimensões e linearmente separáveis:

- Estrelas
- Bolinhas

Como definir a região de decisão?

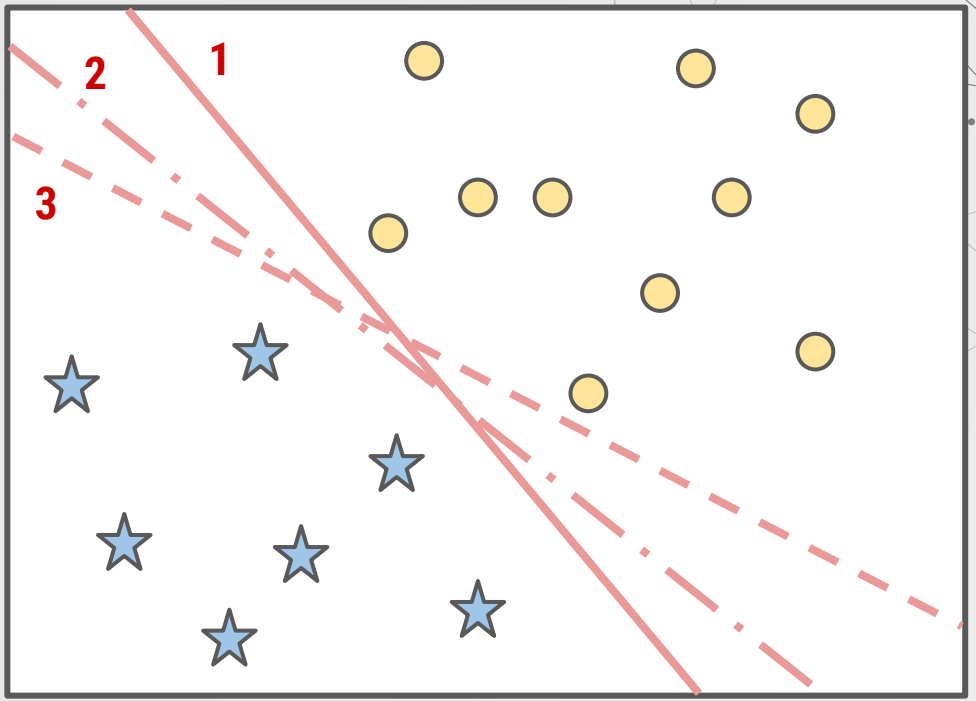


Margem

Por simplicidade, vamos supor duas classes em um espaço de duas dimensões e linearmente separáveis:

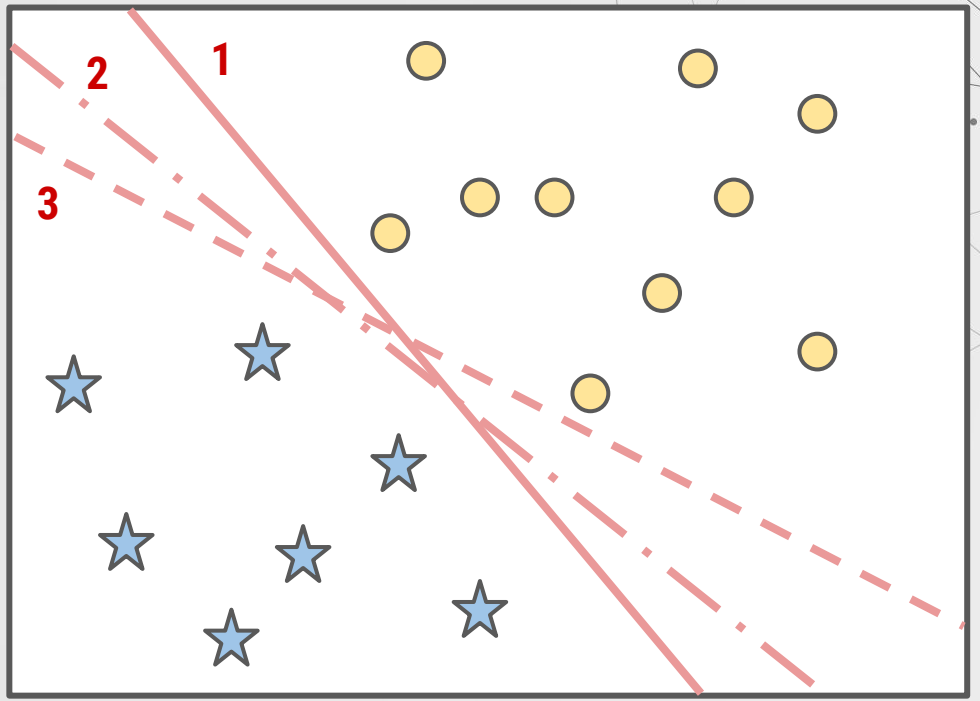
- Estrelas
- Bolinhas

Como definir a região de decisão?



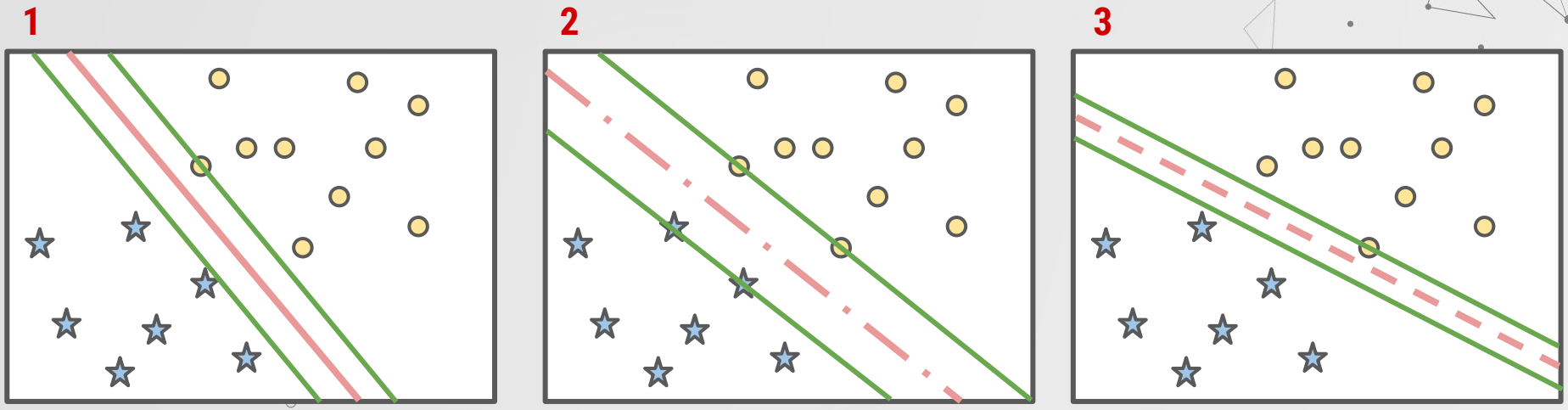
Margem

Como definir a região de decisão?



Como encontrar o melhor hiperplano?

Margem



A margem superior e a margem inferior são paralelas e estão a uma mesma distância do hiperplano

Margem

Como definir a região de decisão?

Como encontrar o melhor hiperplano?

Maximizando a margem.

Este problema possui solução analítica.

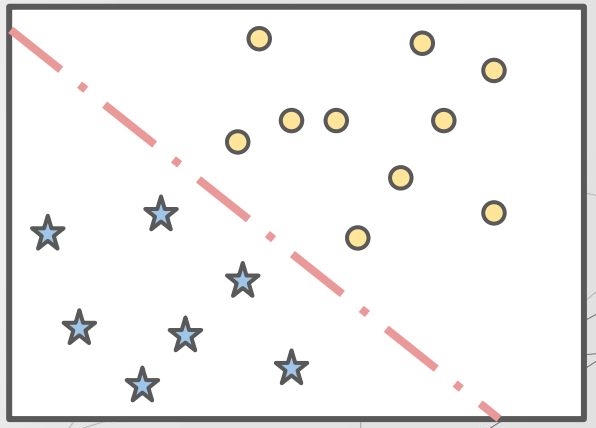
Perceptron (Rosenblatt, 1958)

Uma forma de descrever o problema do hiperplano é **minimizando a distância de pontos classificados erroneamente ao hiperplano**. Suponha y = {-1, 1} as duas respostas/classes possíveis:

- Se $y_i = 1$ foi classificado errado: $x_i^T\beta + \beta_0 < 0$
- Se $y_i = -1$ foi classificado errado: $x_i^T\beta + \beta_0 > 0$

Ou seja, deve-se minimizar:

$$D(\beta, \beta_0) = -\sum_{i \in \mathcal{M}} y_i (x_i^T \beta + \beta_0)$$



Support Vector Machines

No caso do SVM, queremos maximizar a distância M, ou seja:

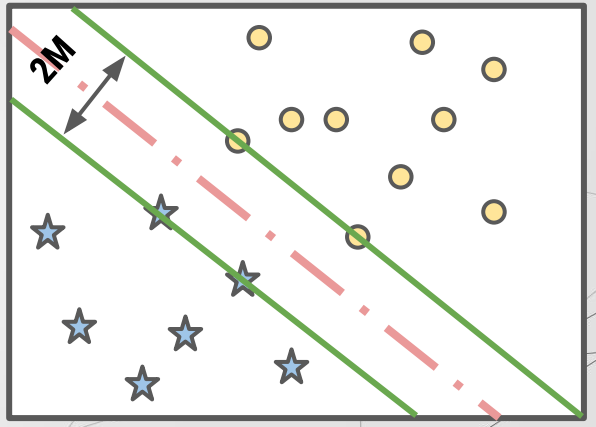
$$\max_{\beta,\beta_0,||\beta||=1} M$$

Sujeito à (i=1,...,N):

$$y_i(x_i^T\beta + \beta_0) \geq M$$

Ou ainda...

$$\max_{\beta,\beta_0} M \qquad \frac{1}{||\beta||} y_i(x_i^T\beta + \beta_0) \geq M$$



Podemos, arbitrariamente, definir:

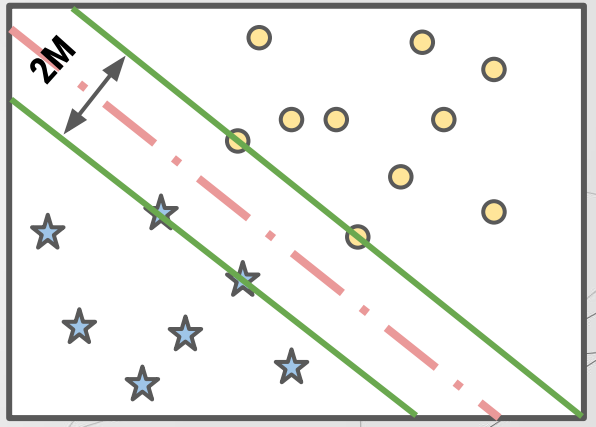
$$||\beta|| = \frac{1}{M}$$

Essa definição manterá a desigualdade mostrada no slide anterior e facilitará a otimização do algoritmo. Ou seja, um problema equivalente é:

$$\min_{\beta,\beta_0} \frac{1}{2}||\beta||^2$$

Sujeito à:

$$y_i(x_i^T\beta + \beta_0) \geq 1$$



Formulando com multiplicadores de Lagrange:

$$L_P = \frac{1}{2}||\beta||^2 - \sum_{i=1}^{N} \alpha_i [y_i (x_i^T \beta + \beta_0) - 1]$$

Fazendo as derivadas iguais a zero e substituindo em Lp, temos:

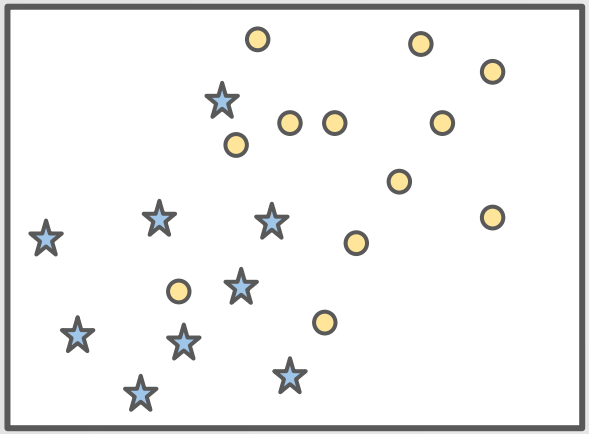
$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k$$

A solução é obtida maximizando L_D e esta solução deve ainda satisfazer, para todo i:

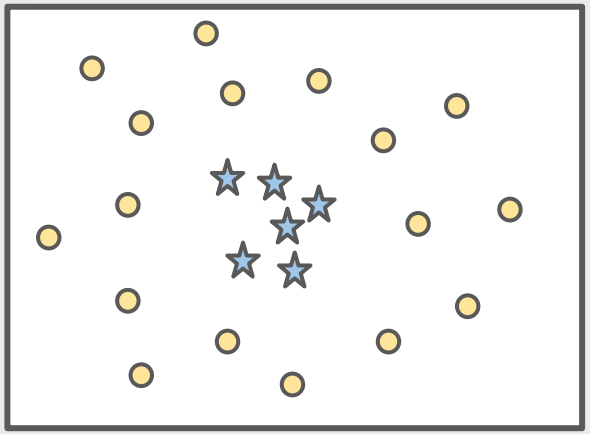
$$\alpha_i [y_i (x_i^T \beta + \beta_0) - 1] = 0$$

Situações comumente encontradas na vida real e que ignoramos até agora

Região de confusão



Dados não-linearmente separáveis



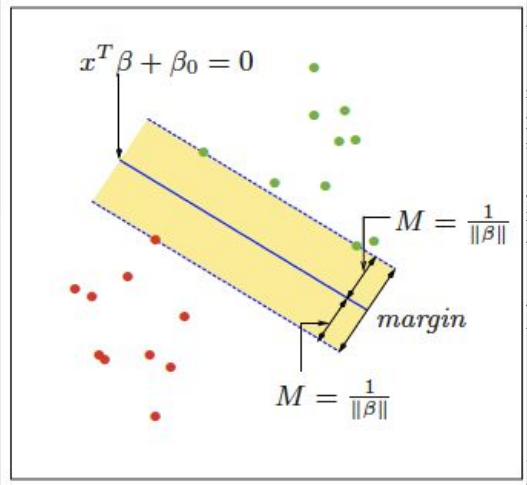
Modificação para lidar com a região de confusão:

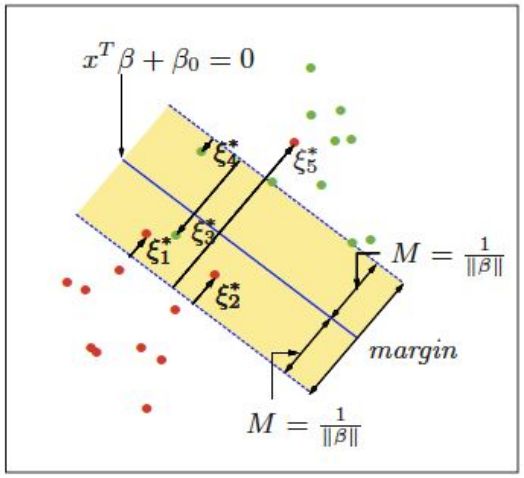
$$\min_{\beta,\beta_0} \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{N} \xi_i$$

Sujeito à:

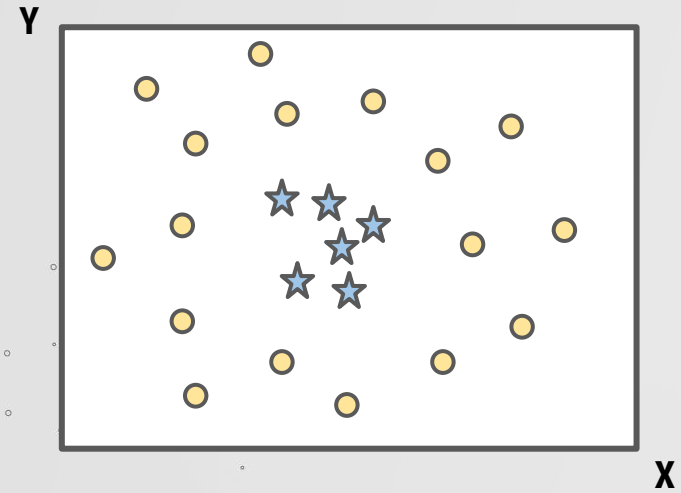
$$\xi_i \geq 0, y_i(x_i^t\beta + \beta_0) \geq 1 - \xi_i, \forall i$$

C é um “custo” e é um hiperparâmetro do modelo. Para o caso mostrado anteriormente sem região de confusão: C = infinito.

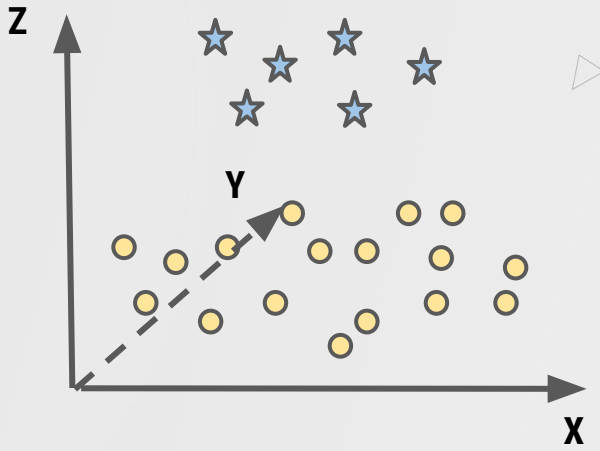




Para lidar com dados não-linearmente separáveis, utilizamos o que chamamos de **Kernel tricks**.



$$z = x^2 + y^2$$



Para lidar com dados não-linearmente separáveis, utilizamos o que chamamos de **Kernel tricks**.

Polinomial grau n: $K(x, x') = (1 + < x, x' >)^d$

Função de base radial (RBF): $K(x, x') = exp(-\gamma||(x - x'||^2)$

Exemplo
Classificação de estrelas e galáxias

Alguns trabalhos publicados

- Self-supervised Learning for Astronomical Image Classification - Martinazzo, Espadoto & Hirata 2020
- Star-Galaxy Separation via Gaussian Processes with Model Reduction - Goumiri et al. 2020 (submetido)
- The miniJPAS survey: star-galaxy classification using machine learning - Baqui et al. 2020 (submetido)
- Unsupervised Star Galaxy Classification with Cascade Variational Auto-Encoder - Sun et al. 2019
- Unsupervised star, galaxy, qso classification: Application of HDBSCAN - Logan & Fotopoulou 2019

Exemplo
Classificação de estrelas e galáxias

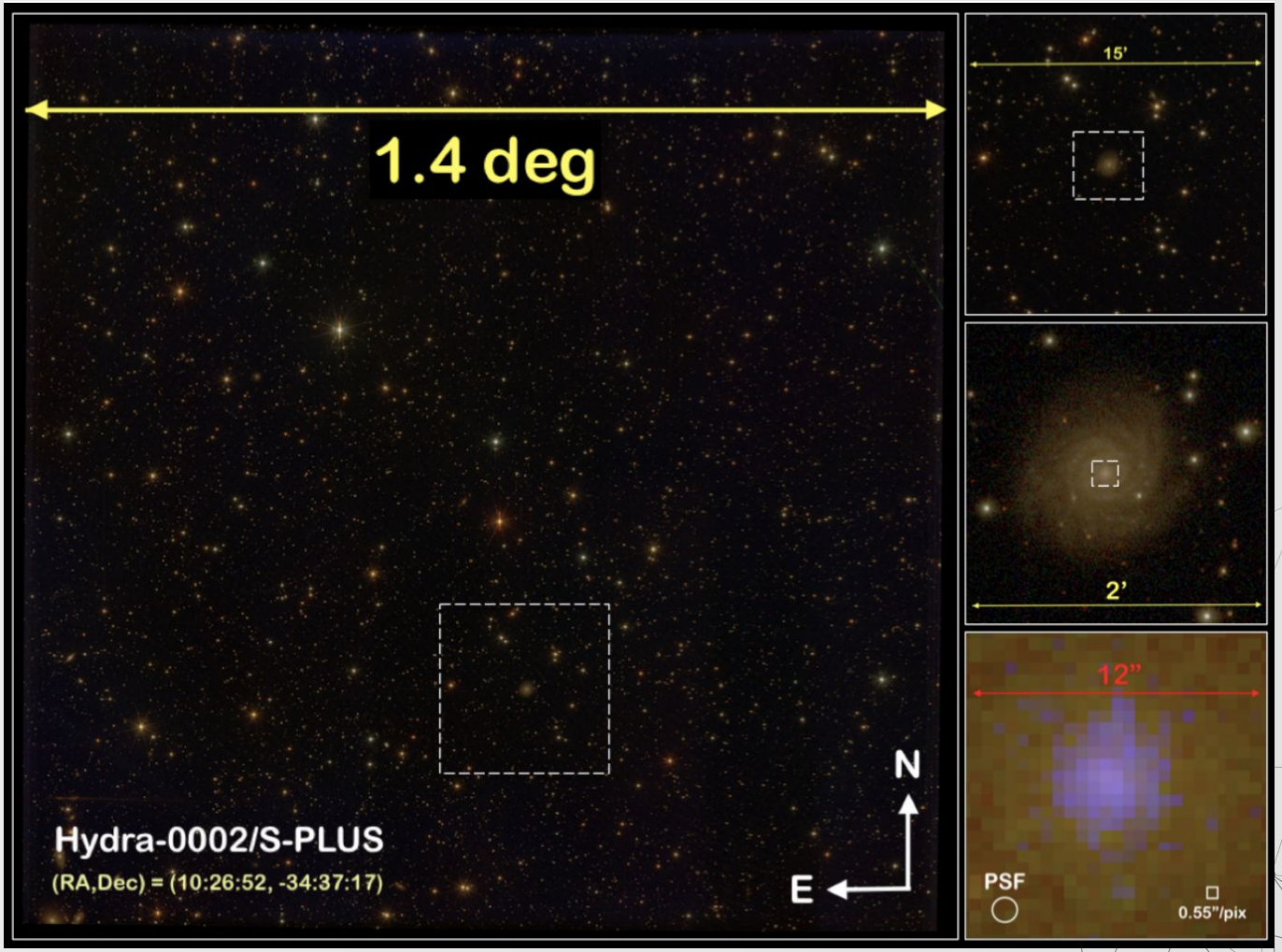
SUPER INTERESSANTE

Ciência

Astrônomos brasileiros estão mapeando metade do céu no Hemisfério Sul

Projeto S-PLUS, que está em operação no Chile, é o maior levantamento do céu já realizado pela comunidade astronômica brasileira.

Por **A. J. Oliveira** Atualizado em 10 jul 2019, 18h43 - Publicado em 9 jul 2019, 15h46

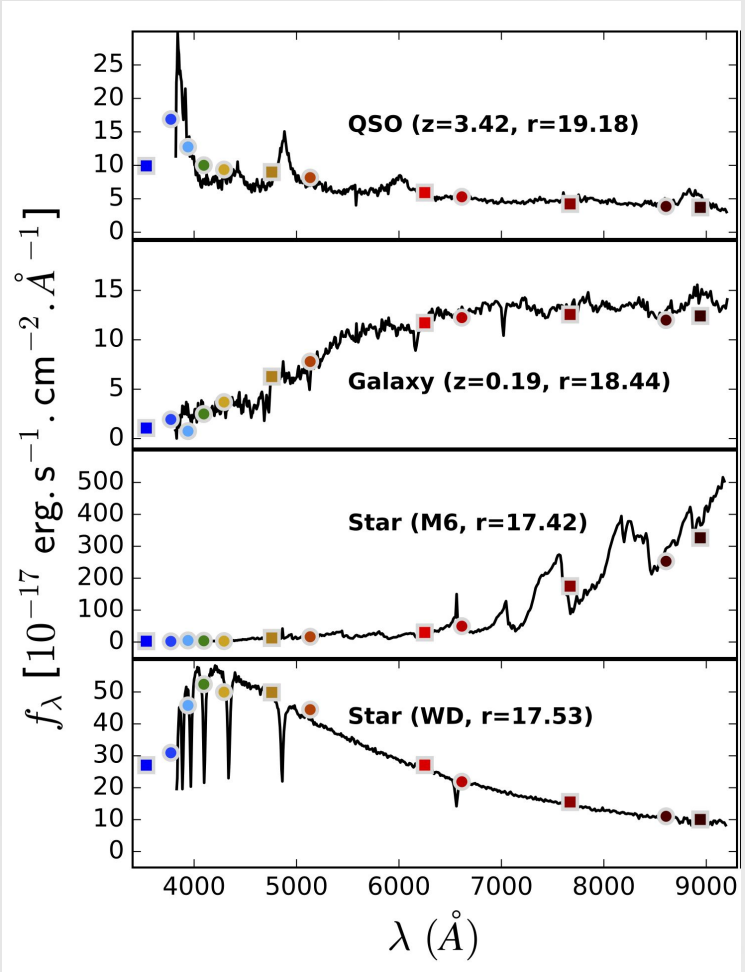

1.4 deg
Hydra-0002/S-PLUS
(RA,Dec) = (10:26:52, -34:37:17)
N
E
15'
2'
12"
PSF
0.55"/pix

On the discovery of stars, quasars, and galaxies in the Southern Hemisphere with S-PLUS DR1

L. M. I. Nakazono[1]★, C. Mendes de Oliveira[1], N. S. T. Hirata[2], S. Jeram[3]
C. Queiroz[4], Stephen S. Eikenberry[3], A. H. Gonzalez[3], R. Abramo[4], R. Overzier[5],
M. Espadoto[2], A. Martinazzo[2], L. Sampedro[1], F. R. Herpich[1], A. Cortesi[1],
F. Almeira-Fernandes[1], A. Werle[1], C. E. Barbosa[1], L. Sodré Jr.[1], E. V. Lima[1],
M. L. Buzzo[1], K. Menéndez-Delmestre[6], S. Akras[7], Alvaro Alvarez-Candal[5,8,9],
A. R. Lopes[5], E. Telles[5], W. Schoenell[10], A. Kanaan[11], T. Ribeiro[12]

[1] *Instituto de Astronomia, Geofísica e Ciências Atmosféricas da U. de São Paulo, Cidade Universitária, 05508-900, São Paulo, SP, Brazil*
[2] *Departamento de Ciência da Computação, Instituto de Matemática e Estatística da USP, Cidade Universitária, 05508-090, São Paulo, SP, Brazil*
[3] *Department of Astronomy, University of Florida, 211 Bryant Space Center, Gainesville, FL 32611, USA*
[4] *Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, SP, Rua do Matão 1371, São Paulo, Brazil*
[5] *Observatório Nacional / MCTIC, Rua General José Cristino 77, Rio de Janeiro, RJ, 20921-400, Brazil*
The remaining institutions are at the end of the paper.

Dados públicos: http://www.splus.iag.usp.br/data/

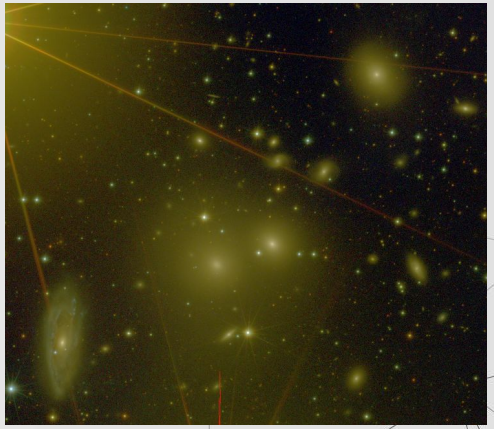


Nakazono et al. 2020 (submetido)

Features	Algorithm Training time (s)	Class	Precision (P)	Recall (R)	F-measure (F)
(i) 12 S-PLUS bands: uJAVA, J0378, J0395, J0410, J0430, J0515, J0660, J0861, g, r, i and z	(a) SVM 821 ± 85	QSO	0.6219 ± 0.0125	0.9031 ± 0.0065	0.7365 ± 0.0083
		STAR	0.6618 ± 0.0037	0.723 ± 0.0027	0.691 ± 0.0026
		GAL	0.8032 ± 0.0022	0.6983 ± 0.0052	0.7471 ± 0.0039
		Macro-averaged F-measure ($\bar{F}$): 0.7249 ± 0.0043			
	(b) RF 30.3 ± 0.9	QSO	0.9093 ± 0.0083	0.8735 ± 0.0076	0.891 ± 0.0072
		STAR	0.9588 ± 0.0019	0.9026 ± 0.002	0.9299 ± 0.0012
		GAL	0.9248 ± 0.0012	0.9647 ± 0.0003	0.9443 ± 0.0007
		Macro-averaged F-measure ($\bar{F}$): 0.9218 ± 0.0024			
(ii) 12 S-PLUS bands + 2 WISE bands: uJAVA, J0378, J0395, J0410, J0430, J0515, J0660, J0861, g, r, i, z, W1 and W2	(a) SVM 694 ± 104	QSO	0.7932 ± 0.0153	0.9304 ± 0.0055	0.8562 ± 0.0078
		STAR	0.8753 ± 0.0031	0.8784 ± 0.0046	0.8768 ± 0.0019
		GAL	0.9188 ± 0.0025	0.8912 ± 0.0038	0.9048 ± 0.0024
		Macro-averaged F-measure ($\bar{F}$): 0.8793 ± 0.0037			
	(b) RF 30.9 ± 1.7	QSO	**0.9539 ± 0.0036**	0.9391 ± 0.0073	0.9464 ± 0.0048
		STAR	**0.9713 ± 0.0014**	**0.9679 ± 0.0009**	**0.9696 ± 0.0002**
		GAL	**0.9715 ± 0.0008**	**0.9760 ± 0.0006**	**0.9737 ± 0.0006**
		Macro-averaged F-measure ($\bar{F}$): **0.9633 ± 0.0018**			

Dados do exemplo em R

- 5k galáxias e 5k estrelas (amostradas aleatoriamente de um conjunto maior para fins didáticos)
- Sem dados faltantes (também para fins didáticos)
- Variáveis preditoras:
  - Largura à meia altura do perfil de brilho (FWHM)
  - Semieixo maior
  - Semieixo menor
  - Raio de Kron
  - Magnitude em cada uma das 12 bandas do S-PLUS



Código disponível no GitHub: https://github.com/marixko/stargalaxy_SVM

Table 2: From a qualitative examination of a sample of ∼ 200 refereed publications from 2017 to February 2019, a mapping emerges between the nature of astronomical data and the types of machine learning and artificial intelligence algorithms that are being applied. The table presents a summary of the types of astronomical data and the algorithms that appeared most regularly. The purpose of the table is to provide a convenient starting point for selecting an algorithm that has been used successfully for each data type.

Data/Method	ANN	CNN	GAN	SVM	DT	RF	DBSCAN	k-NN	k-M
Image	•	•	•	•	•	•		•	
Spectroscopy	•	•		•		•			•
Photometry	•				•	•	•		•
Light curve		•				•			
Time Series	•	•			•	•	•		
Catalogue	•			•	•	•	•	•	
Simulation	•	•	•	•		•			

ANN = Artificial Neural Network; CNN = Convolutional Neural Network; GAN = Generative Adversarial Network; SVM = Support Vector Machine; DT = Decision Tree; RF = Random Forest; DBSCAN = Density-based spatial clustering of applications with noise; k-NN = k-Nearest Neighbours; k-M = k-means clustering

Surveying the reach and maturity of machine learning and artificial intelligence in astronomy - Fluke & Jacobs 2019