

Support Vector Machines (**SVM**)

Ruan Heleno Correa da Silva

Sumário

- Introdução
- Support Vector Machines (SVM)
- Vetores de Suporte e Hiperplano
- Execução
- Hiperplano para dados não lineares
- Prós e Contras
- Aplicação
- Referências



Introdução

→ Modelos de Classificação

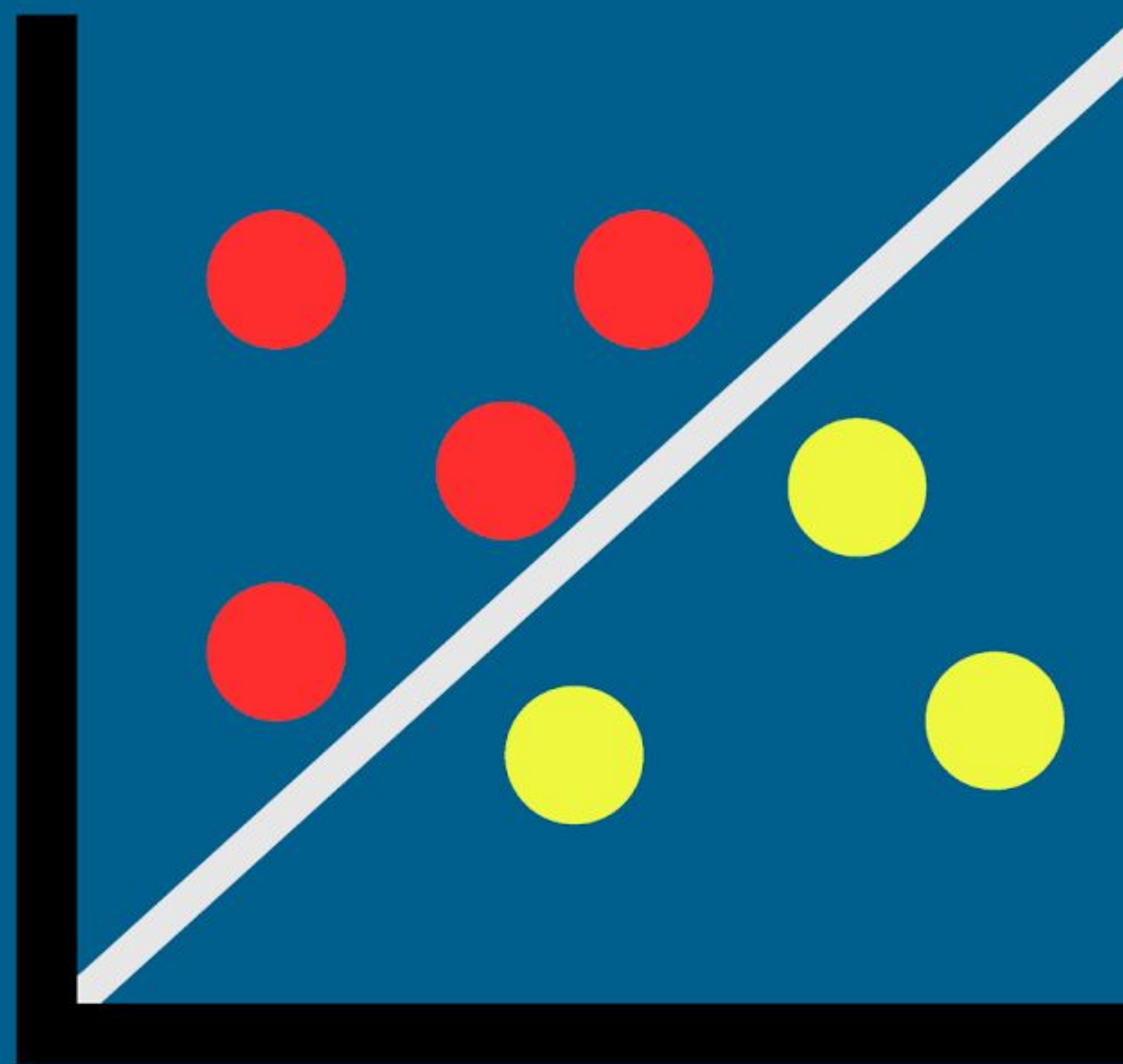
- Árvore de Decisão
- Regressão
 - Linear
 - Logística
- KNN
- Support Vector Machines (SVM)
 - Linear
 - Não-Linear



Support Vector Machines (SVM)

- Idealizado em 1979 por Vladimir Vapnik e alguns colegas
- Métodos de ML de análise de dados e reconhecimento de padrões
- Aprendizado de máquina supervisionado
- Treinamento e classificação de um dataset
 - ◆ classificador linear binário não probabilístico



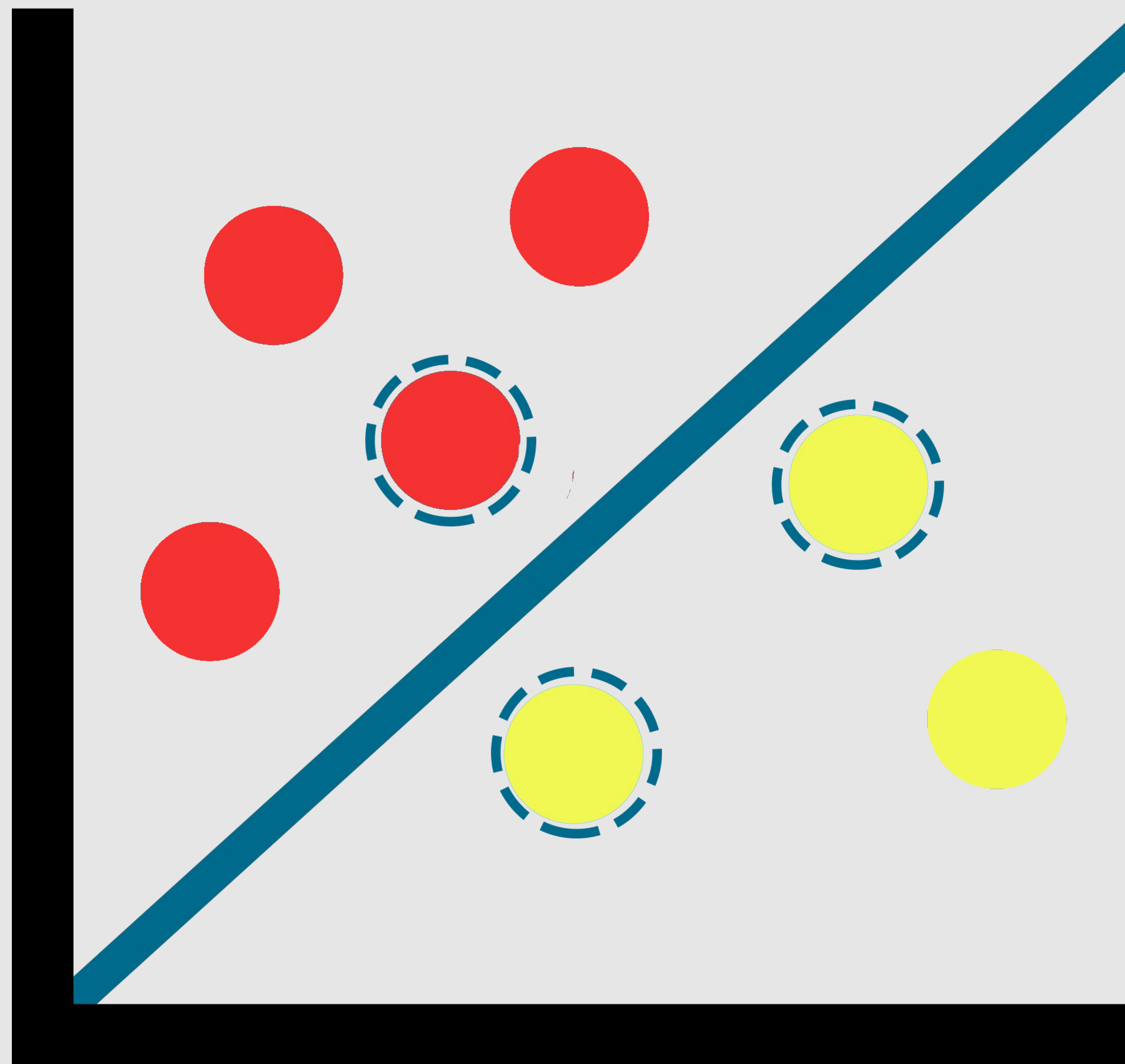


Vetores de Suporte e Hiperplano



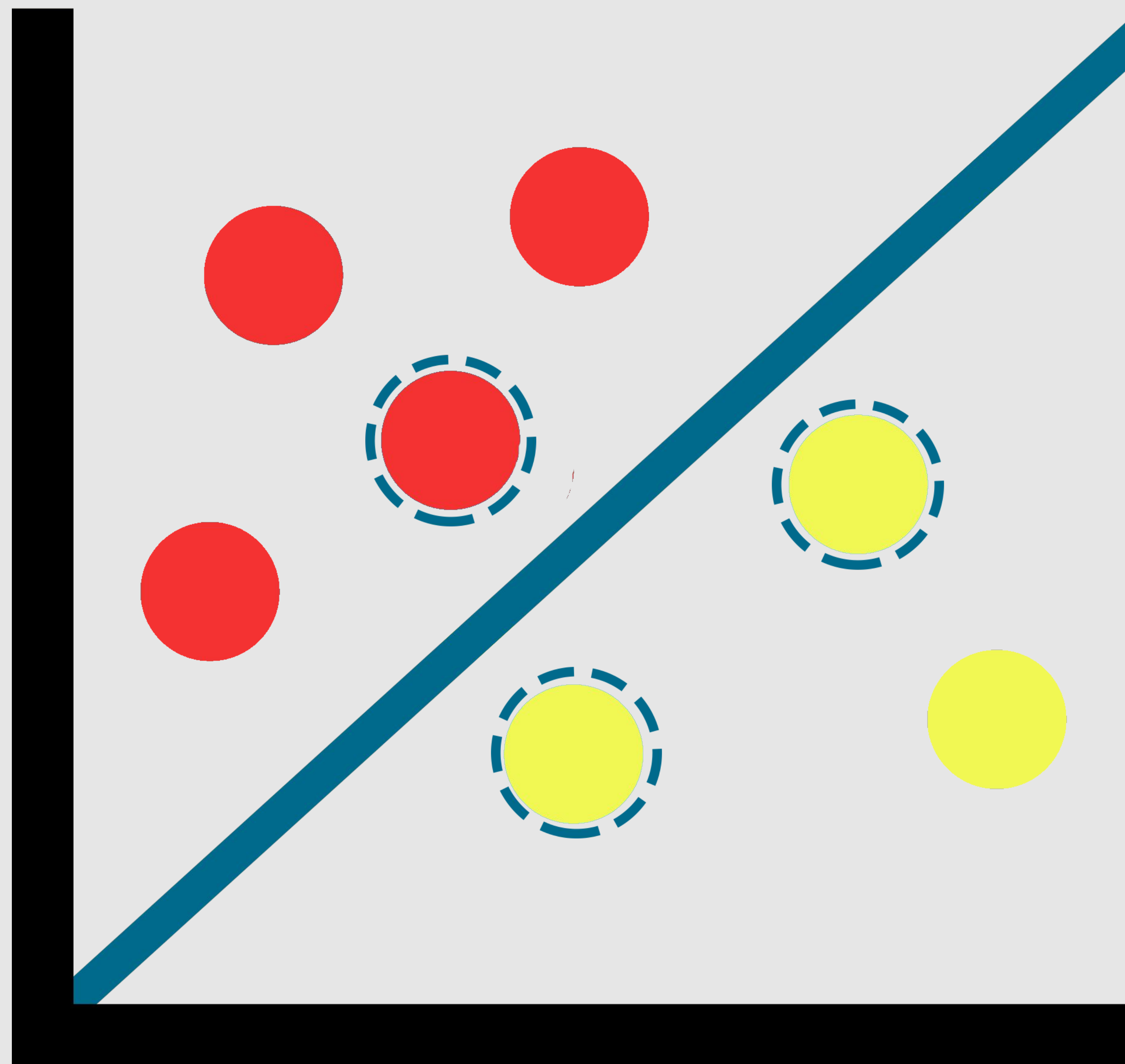
Vetores de Suporte e Hiperplano

- Vetores de suporte são as coordenadas da observação individual.
- São vetores de pontos usados para formar a margem que cria o limite entre as classes



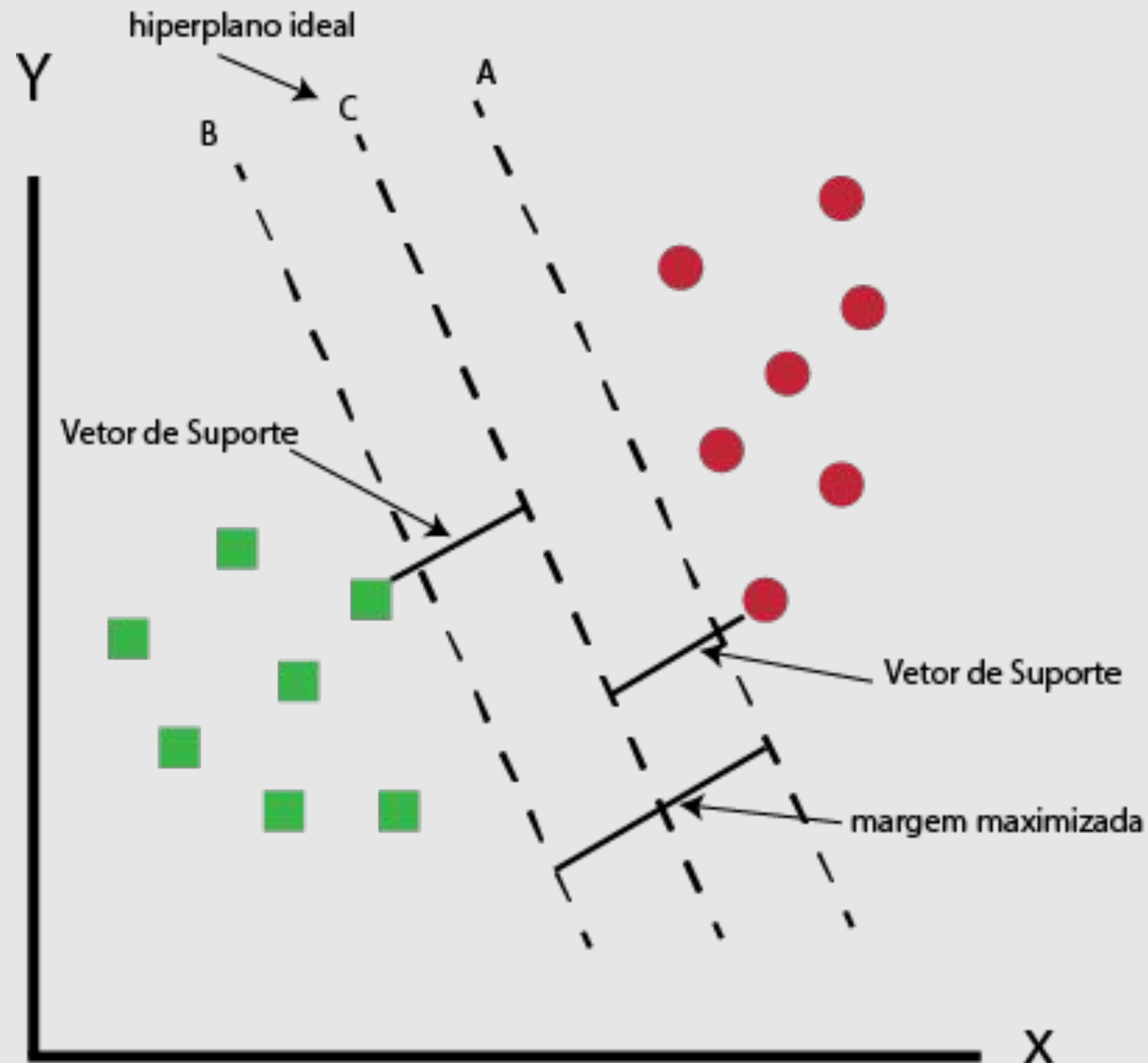
Vetores de Suporte e Hiperplano

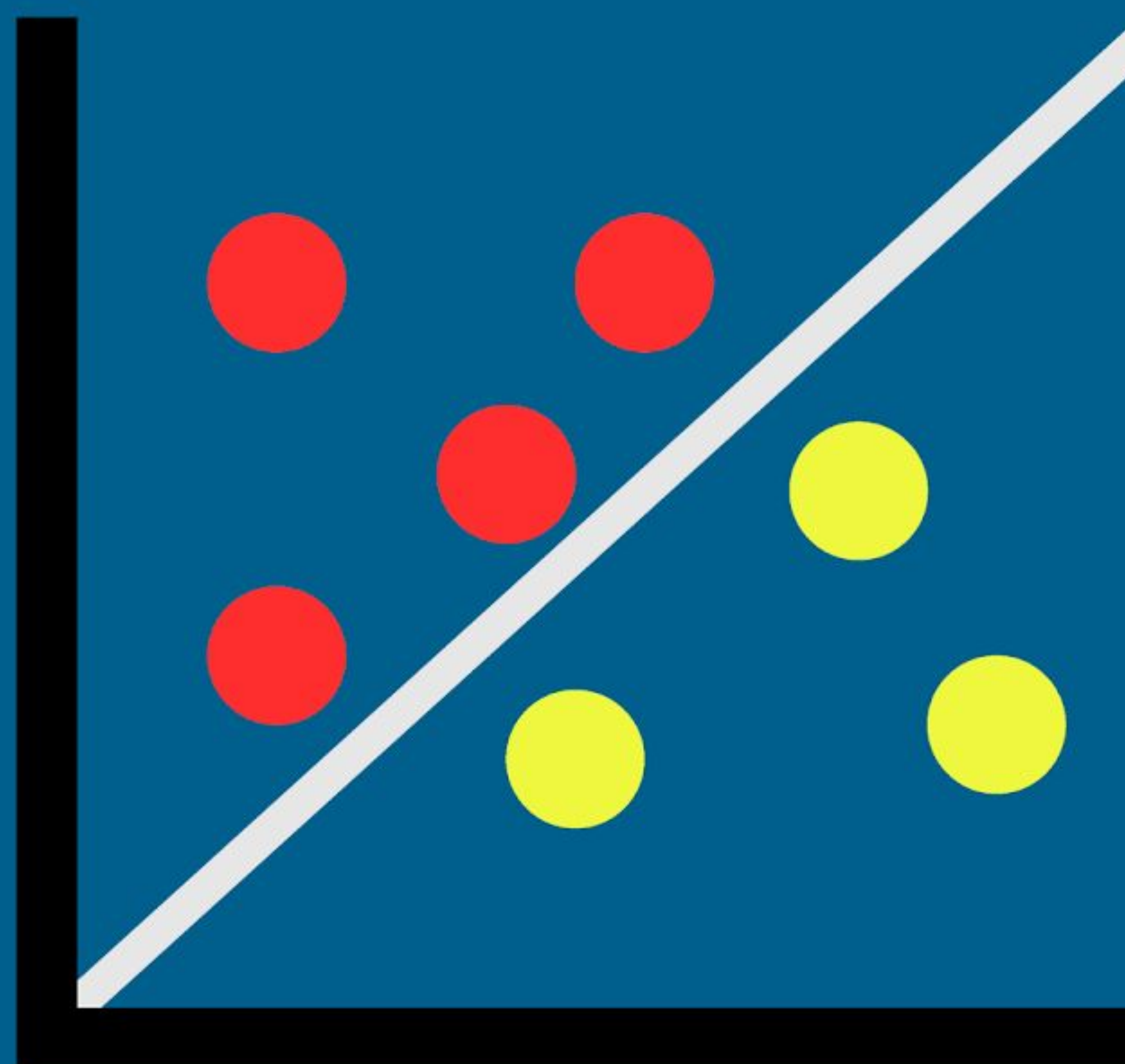
- Hiperplano é uma linha/fronteira que melhor se agrega entre tais vetores de suporte
- Podemos ter diversos hiperplanos possíveis



Vetores de Suporte e Hiperplano

→ O SVM sempre tenta encontrar o hiperplano que vai maximizar a margem entre as classes





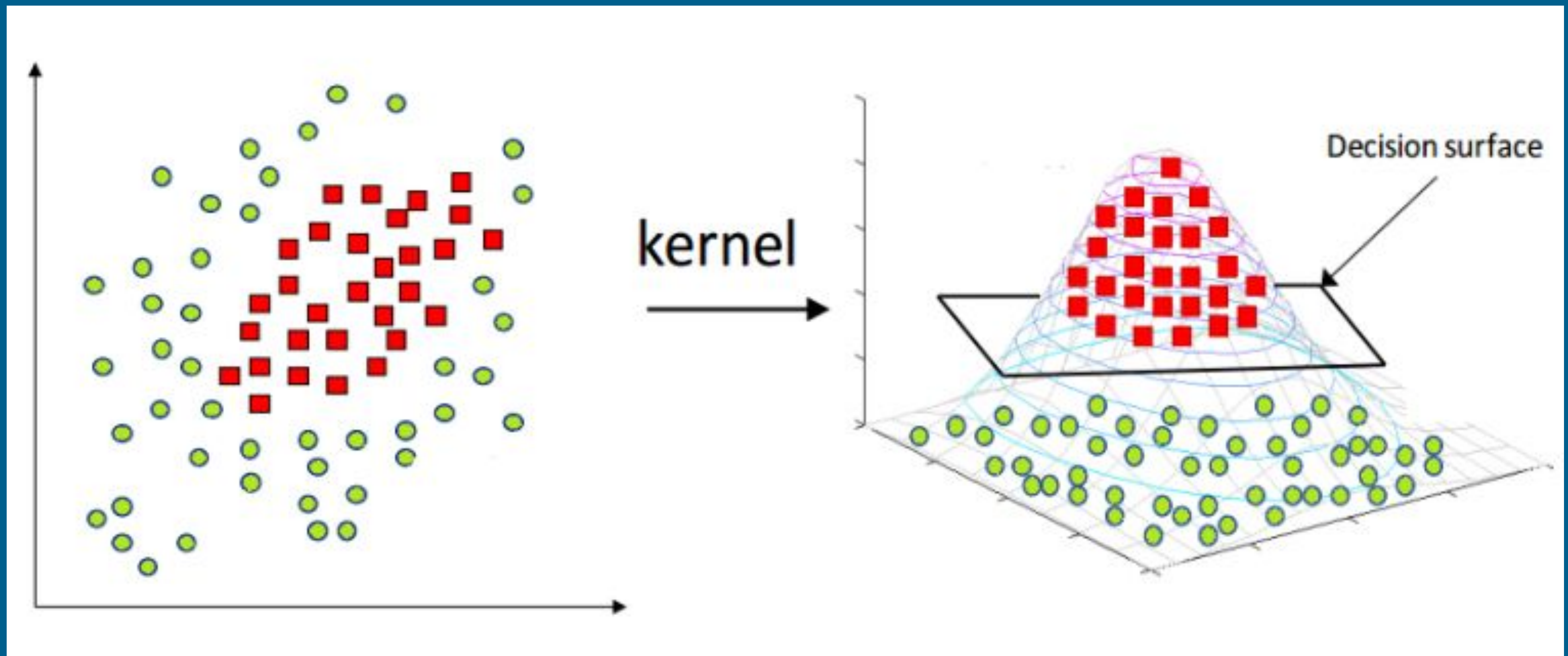
Execução



Execução

- Dado um conjunto de dados de treino, cada um marcado com uma classe pertencendo a uma categoria, SVM cria um modelo que assimila novos exemplos em uma das categorias, fazendo uso de um classificador não probabilístico.
- Nos métodos presentes, plotamos cada item de dados, como um ponto no espaço n -dimensional (onde n é o número de recursos que você tem), com o valor de cada recurso sendo o valor de uma determinada coordenada. Então, nós executamos a classificação encontrando o hiperplano que melhor diferencia as duas classes.





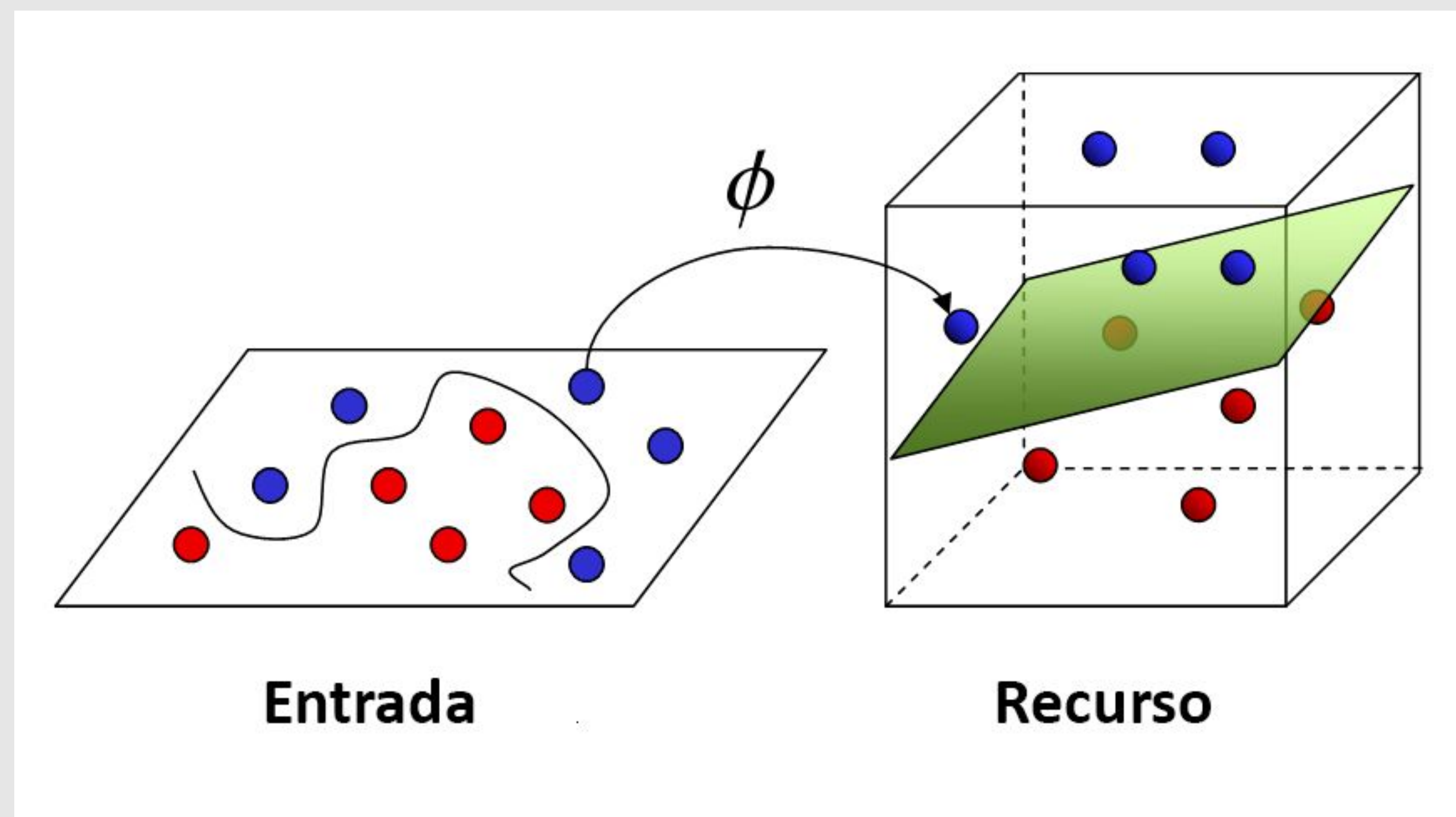
SVM não linear



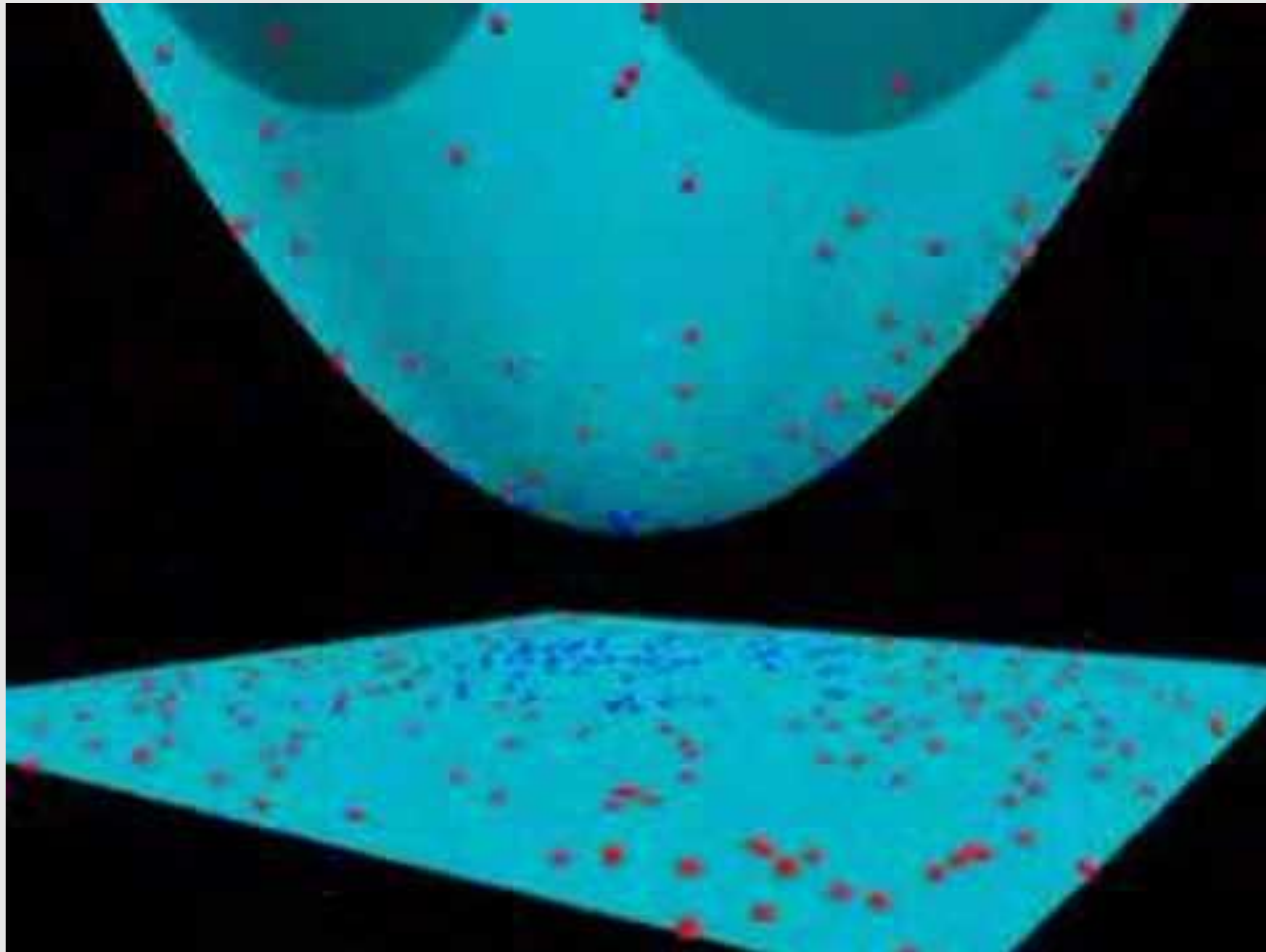
Hiperplano para dados não lineares

→ Kernel Trick

- ◆ se baseia em aplicar uma transformação nos dados e levá-los para um espaço dimensional superior onde ele possa ser separado linearmente e depois os trazemos de volta aplicando a inversa



Hiperplano para dados não lineares



Demonstração



Prós e Contras do SVM

- I. Funciona muito bem com margem de separação clara.
- II. É eficaz nos casos em que o número de dimensões é maior que o número de amostras.
- III. Ele usa um subconjunto de pontos de treinamento na função de decisão (chamamos de vetores de suporte), portanto, também é eficiente em termos de memória.

- I. Não tem um bom desempenho quando temos um grande conjunto de dados porque o tempo de treinamento necessário é grande.
- II. Ele também não funciona muito bem quando o conjunto de dados tem mais ruído, ou seja, as classes de destino estão sobrepostas.



Aplicação

Conjunto de dados do Câncer de mama de Wisconsin

Classificação binária

[sklearn.datasets.load_breast_cancer — scikit-learn 0.24.1 documentation \(scikit-learn.org\)](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html)



Aplicação

- Iris
- Fazer Bolhas

[SupportVectorMachine_SVM/application at main · ruancorrea/SupportVectorMachine_SVM \(github.com\)](#)



→ Trabalho Escrito

- ◆ Explicação mais detalhada
- ◆ Cálculo matemático
 - Linear
 - Não linear



Referências

COELHO, Lucas. Ciência de Dados: O que é, conceito e definição, 2020. Disponível em: <[Ciência de Dados: O que é, conceito e definição | Blog Cetax](#)>. Acesso em 14 de Janeiro de 2021.

SANTANA, Rodrigo. Café com Código #18: O Famoso SVM, 2017. Disponível em: <[Minerando Dados › Café com Código #18: O Famoso SVM](#)>. Acesso em 14 de Janeiro de 2021.

ADDAN, Daniel. Support Vector Machine, 2019. Disponível em: <[IA07.pdf \(ufpr.br\)](#)>. Acesso em 15 de Janeiro de 2021.

J. G. B. CAMPELLO, Ricardo; FERNANDES DE MELLO, Rodrigo. Support Vector Machine - SVM, 2018. Disponível em: <[Reconhecimento de Padrões \(ufu.br\)](#)>. Acesso em 15 de Janeiro de 2021.

LEARN, Scikit. 1.4. Support Vector Machines. Disponível em <[1.4. Support Vector Machines — scikit-learn 0.24.1 documentation \(scikit-learn.org\)](#)>. Acesso em 16 de Janeiro de 2021.

LEARN, Scikit. sklearn.datasets.make_blobs. Disponível em <[sklearn.datasets.make_blobs — scikit-learn 0.24.1 documentation \(scikit-learn.org\)](#)>. Acesso em 17 de Janeiro de 2021.

LEARN, Scikit. sklearn.datasets.load_breast_cancer. Disponível em <[sklearn.datasets.load_breast_cancer — scikit-learn 0.24.1 documentation \(scikit-learn.org\)](#)>. Acesso em 17 de Janeiro de 2021.

MAYERS, Gabriel. Classifying Malignant or Benignant Breast Cancer using SVM, 2020. Disponível em: <[Classifying Malignant or Benignant Breast Cancer using SVM | by Gabriel Mayers | Analytics Vidhya | Medium](#)>. Acesso em 19 de Janeiro de 2021.



