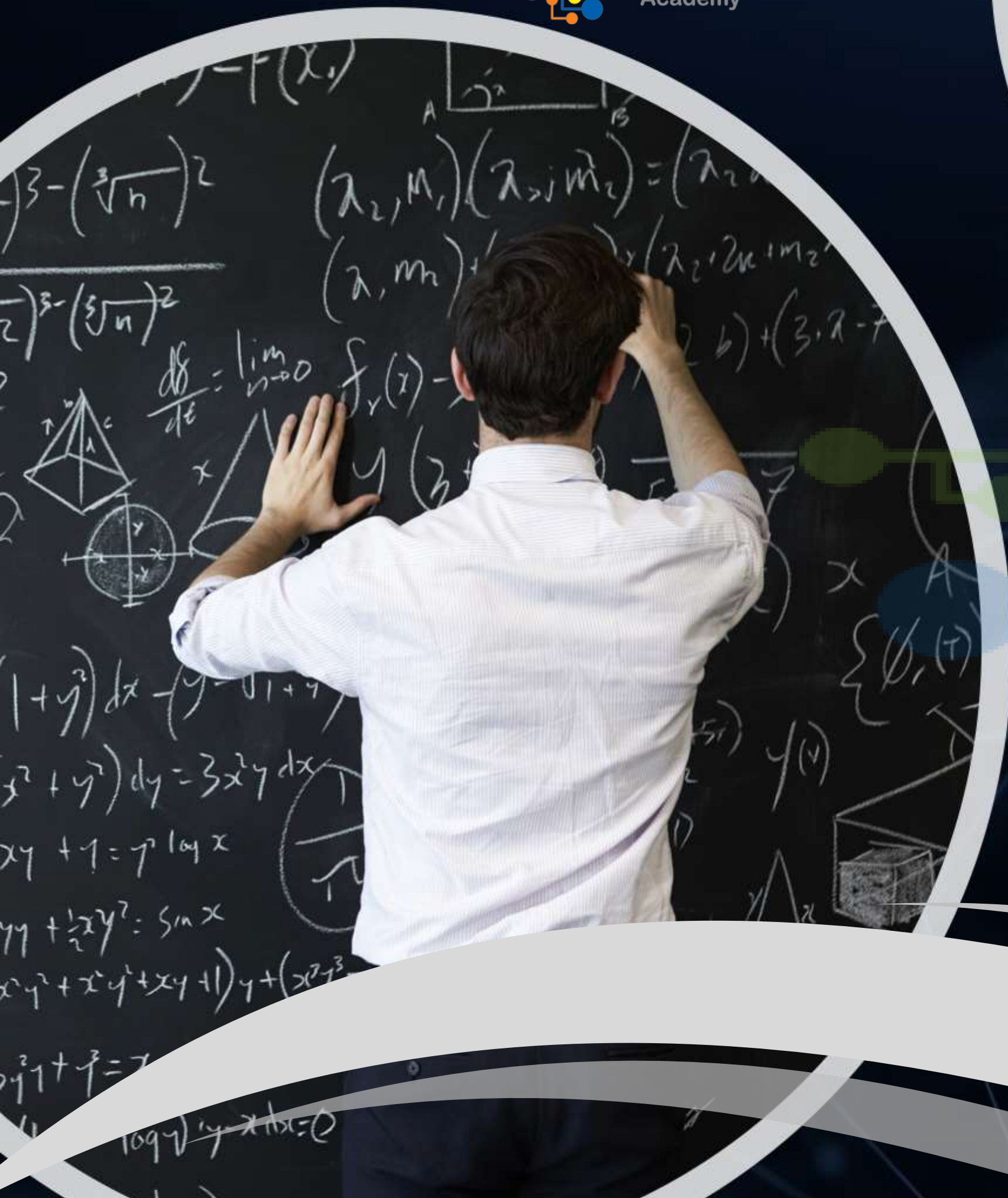




Data Science
Academy

Data Science Academy angelicogfa@gmail.com 5b81f7e45e4cdea2118b4569



Matemática para Machine Learning

A sua base começa aqui!



Matemática para Machine Learning



Laboratórios

Matemática em Modelos de Machine Learning





Data Science
Academy

Data Science Academy angelicogfa@gmail.com 5b81f7e45e4cdea2118b4569

Laboratórios - Matemática em Modelos de Machine Learning





Matemática para Machine Learning



Regressão Linear



Regressão Linear

A regressão linear é uma técnica básica de análise preditiva que usa dados históricos para prever uma variável de saída. É popular para modelagem preditiva porque pode ser facilmente compreendida e explicada.

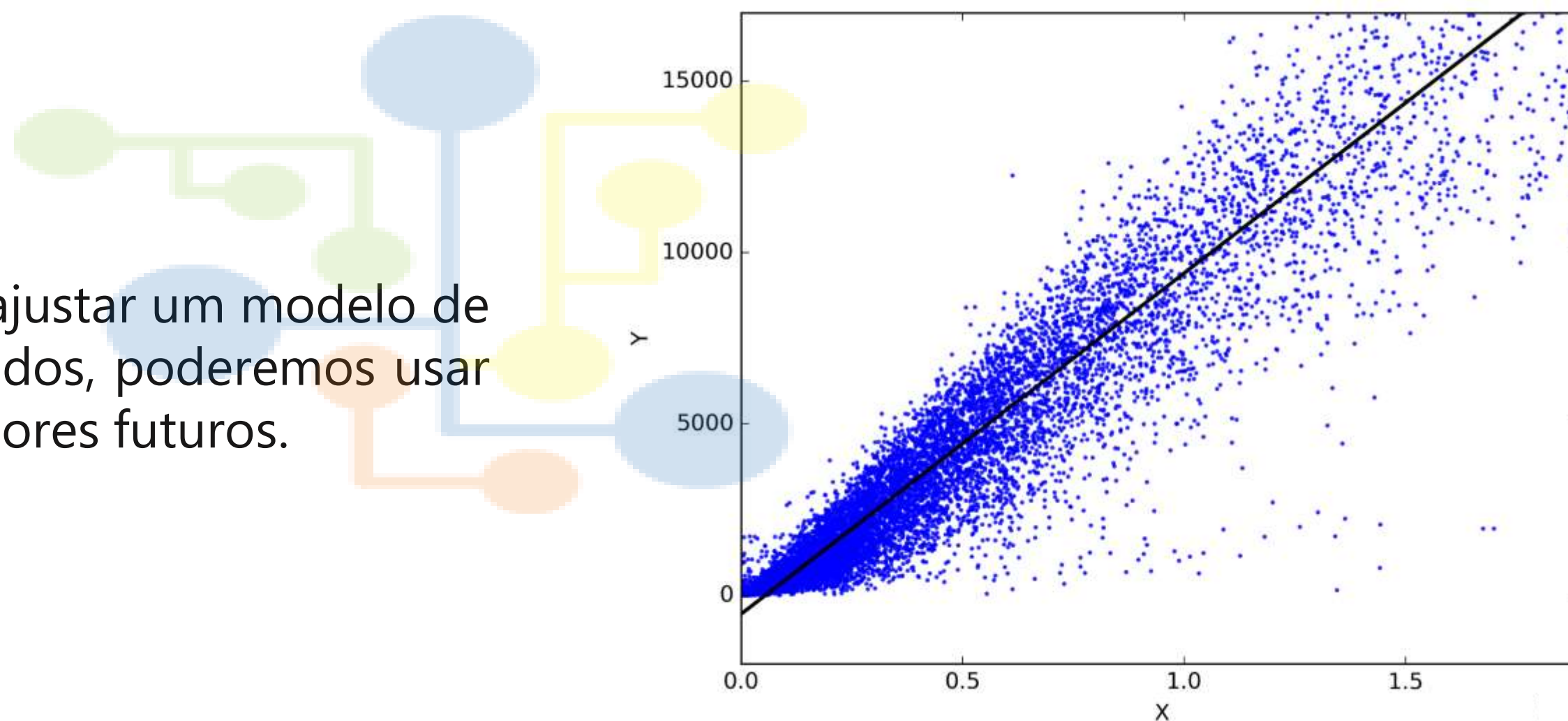
Os modelos de regressão linear têm muitas aplicações do mundo real em uma série de setores como economia (por exemplo, previsão de crescimento), negócios (por exemplo, previsão de vendas de produtos, desempenho de funcionários), ciências sociais (por exemplo, previsão de inclinações políticas de gênero ou raça), saúde (prever os níveis de pressão arterial a partir do peso, início da doença a partir de fatores biológicos) e muito mais.

Compreender como implementar modelos de regressão linear ajuda a resolver problemas importantes, sendo a base de algoritmos mais complexos.



Regressão Linear

A ideia básica é que, se pudermos ajustar um modelo de regressão linear aos dados observados, poderemos usar o modelo para prever quaisquer valores futuros.





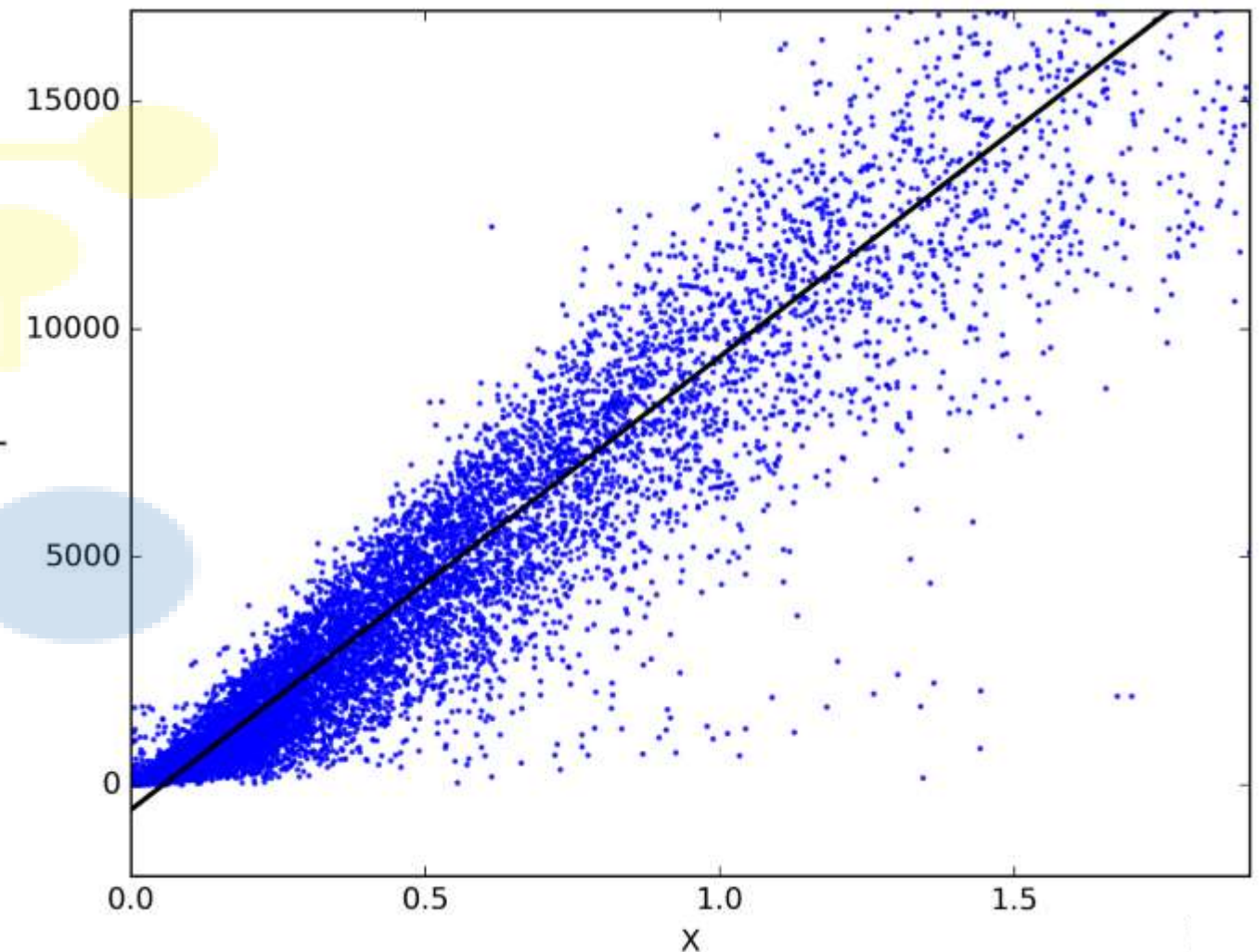
Regressão Linear

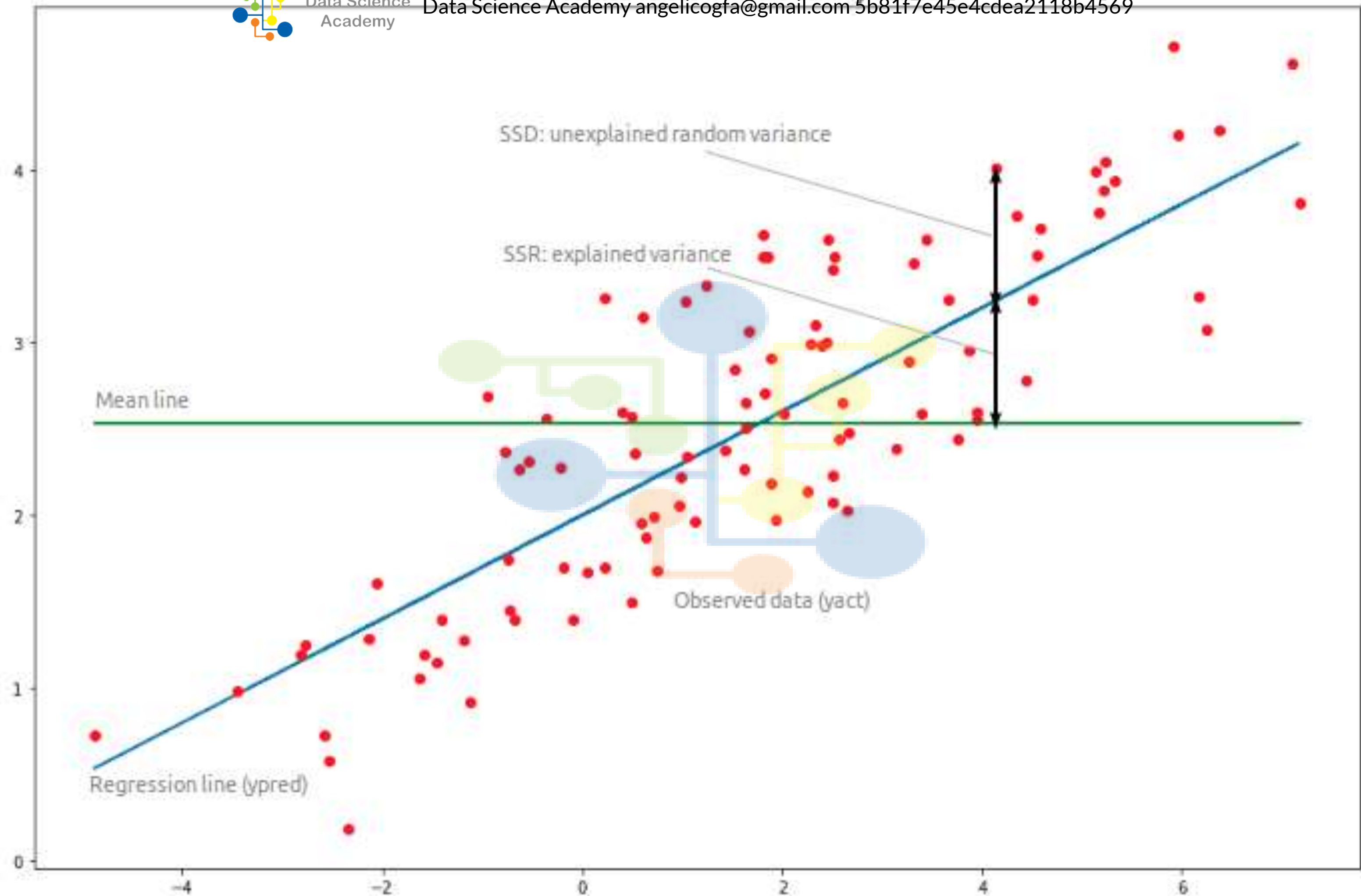
Por exemplo, suponhamos que descobrimos a partir de dados históricos que o preço (P) de uma casa depende linearmente de seu tamanho (S). Na verdade, descobrimos que o preço de uma casa é exatamente 90 vezes seu tamanho. A equação ficará assim:

$$P = 90 * S$$

Com este modelo, podemos então prever o custo de qualquer casa. Se temos uma casa de 1.500 metros quadrados, podemos calcular seu preço como:

$$P = 90 * 1500 = \text{R\$ } 135.000$$







Matemática para Machine Learning



A Matemática dos Algoritmos de Regressão





A Matemática dos Algoritmos de Regressão

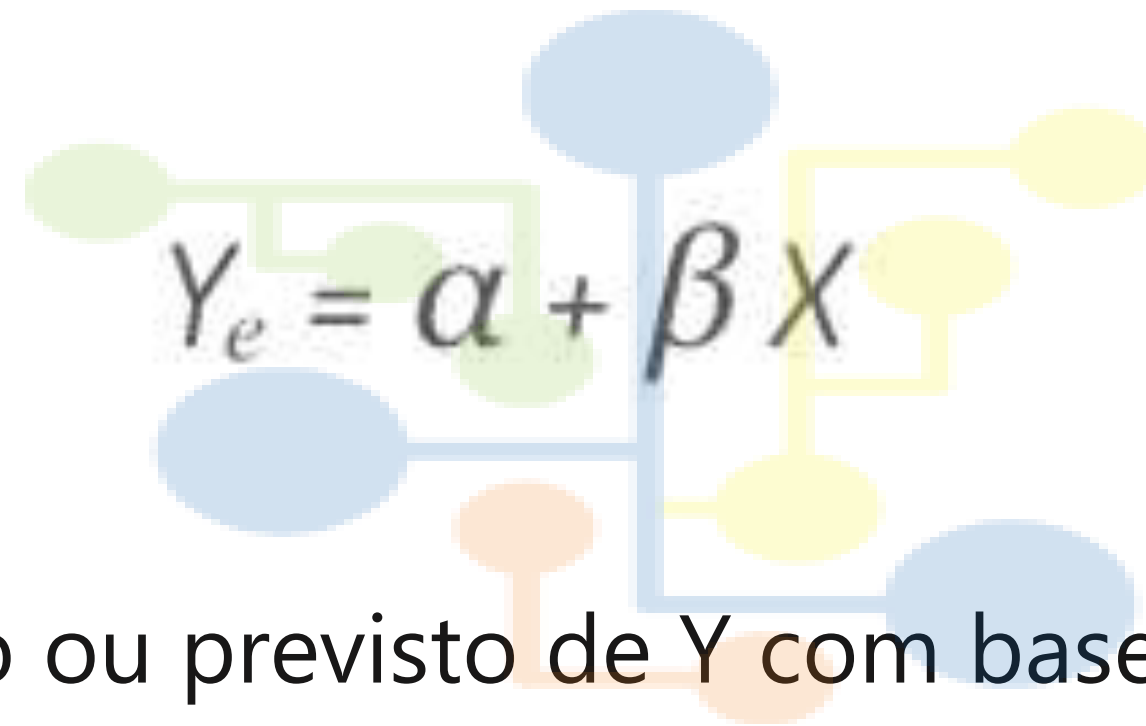
Existem dois tipos de variáveis em um modelo de regressão linear:

- A variável de entrada ou preditor é a variável (ou variáveis) que ajuda a prever o valor da variável de saída. É comumente referido como X .
- A variável de saída é a variável que queremos prever. É comumente referido como Y .



A Matemática dos Algoritmos de Regressão

Para estimar Y usando regressão linear, assumimos a equação:


$$Y_e = \alpha + \beta X$$

onde Y_e é o valor estimado ou previsto de Y com base em nossa equação linear.

Nosso objetivo é encontrar valores estatisticamente significantes dos parâmetros α e β que minimizem a diferença entre Y e Y_e .



A Matemática dos Algoritmos de Regressão

Se formos capazes de determinar os valores ótimos desses dois parâmetros, então teremos a linha de melhor ajuste que podemos usar para prever os valores de Y , dado o valor de X .

Então, como estimamos α e β ? Podemos usar um método chamado mínimos quadrados ordinários (ordinary least squares).



A Matemática dos Algoritmos de Regressão

Ordinary Least Squares

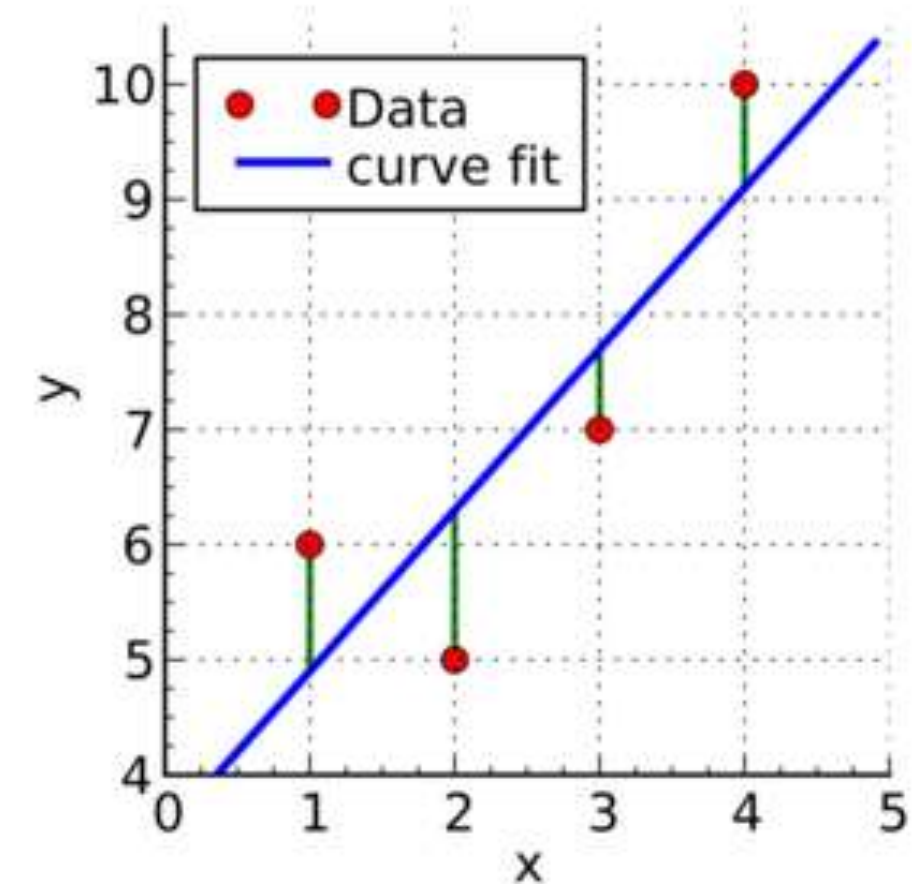
O objetivo do método dos mínimos quadrados é encontrar valores de α e β que minimizem a soma da diferença quadrática entre Y e Y_e . Usando o cálculo podemos mostrar que os valores dos parâmetros desconhecidos são os seguintes:

$$\beta = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\alpha = \bar{Y} - \beta * \bar{X}$$

onde \bar{X} é a média dos valores X e \bar{Y} é a média dos valores Y .

Ou ainda, podemos dizer que β é simplesmente $\text{Cov}(X, Y) / \text{Var}(X)$.





Matemática para Machine Learning

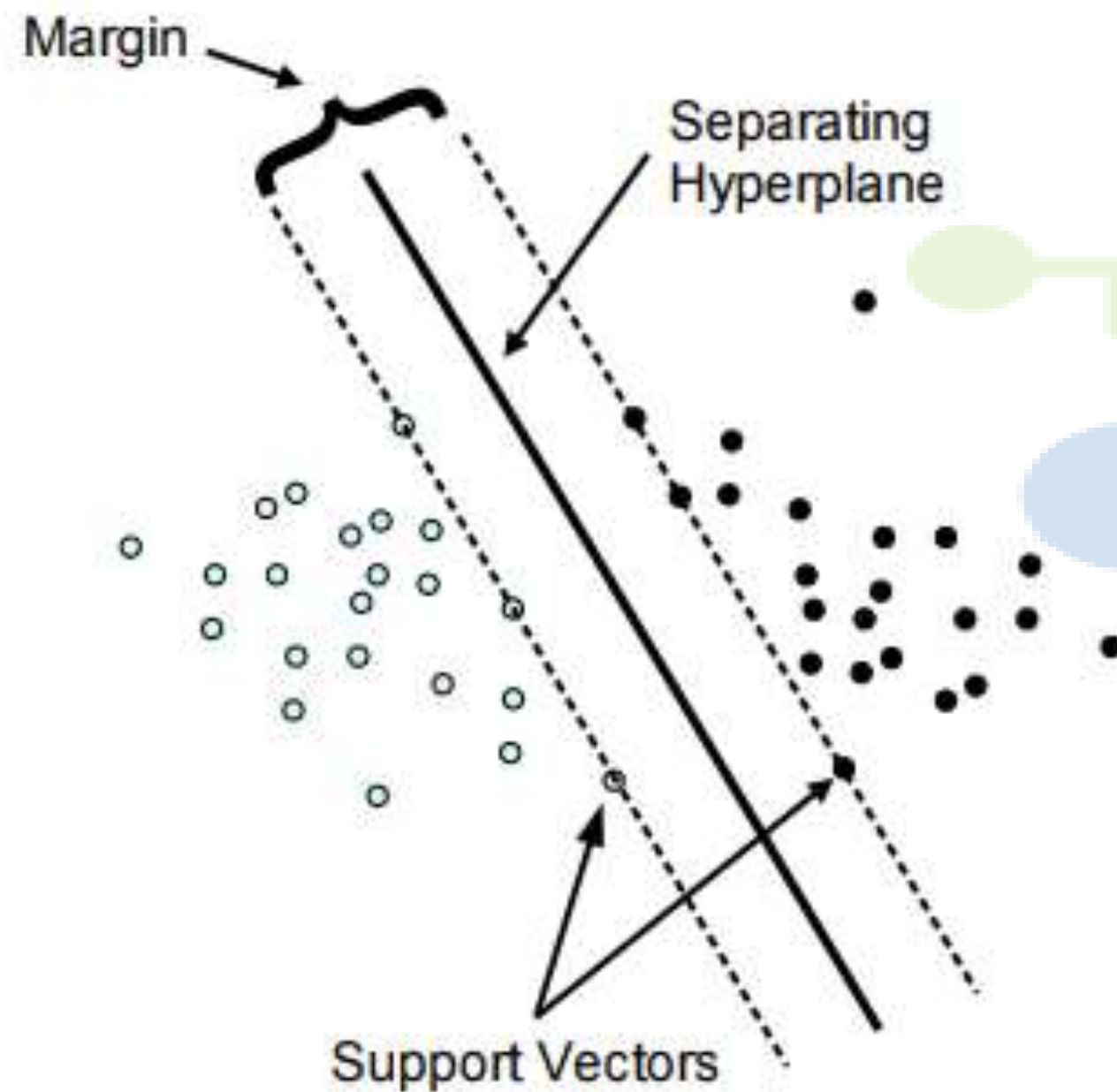


Support Vector Machines





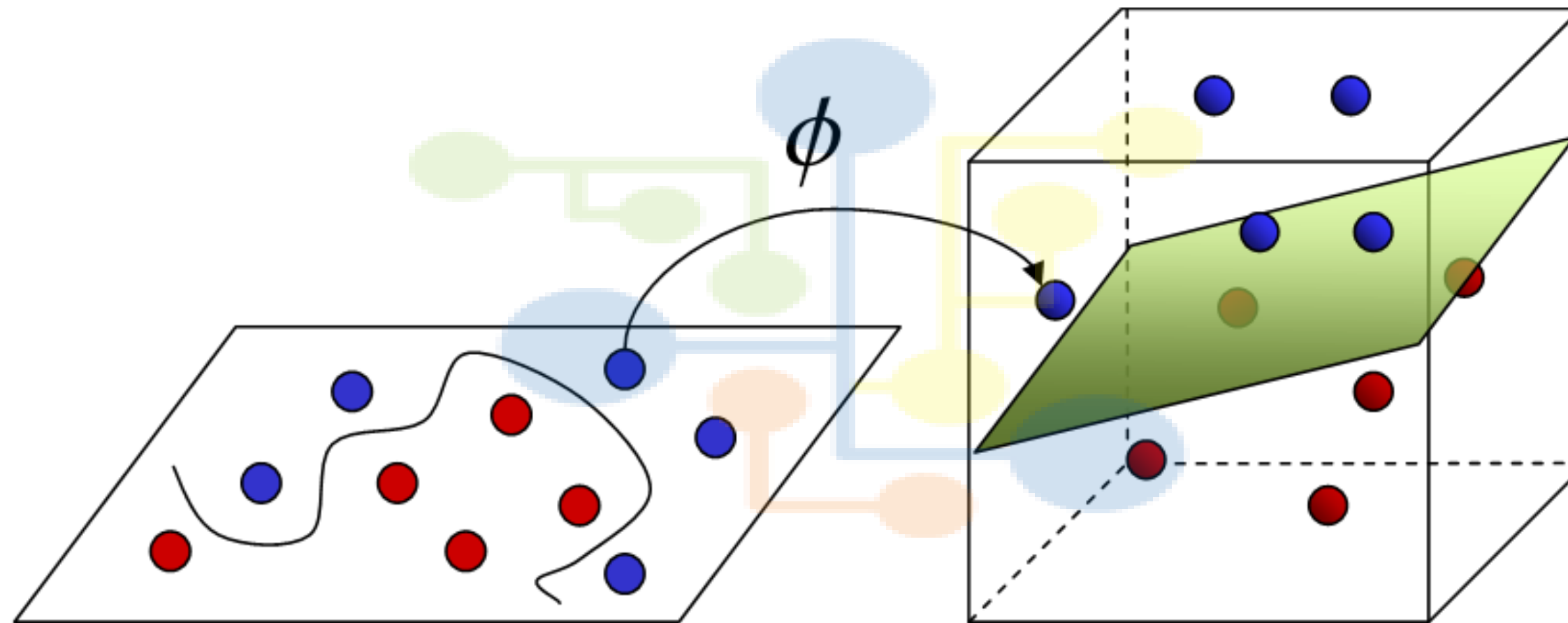
Support Vector Machines



Máquinas de Vetores de Suporte (Support Vector Machines) é um algoritmo de aprendizagem supervisionada para classificação e regressão (e também para detecção de valores outliers), que é bastante versátil, pois pode se ajustar a modelos lineares e não lineares, graças à disponibilidade de funções especiais – as chamadas funções kernel.



Support Vector Machines



Input Space

Feature Space



Matemática para Machine Learning



Principais Características das SVMs





Principais Características das SVMs

- Em caso de *outliers* a SVM busca a melhor forma possível de classificação e, se necessário, desconsidera o *outlier*;
- *É um classificador criado para fornecer separação linear*;
- Funciona muito bem em domínios complicados, em que existe uma clara margem de separação;
- Não funciona bem em conjuntos de dados muito grandes, pois o tempo de treinamento é muito custoso;
- Não funciona bem em conjuntos de dados com grande quantidade de ruídos;
- Se as classes estiverem muito sobrepostas deve-se utilizar apenas evidências independentes.



Matemática para Machine Learning



Hiperplano e Vetores de Suporte



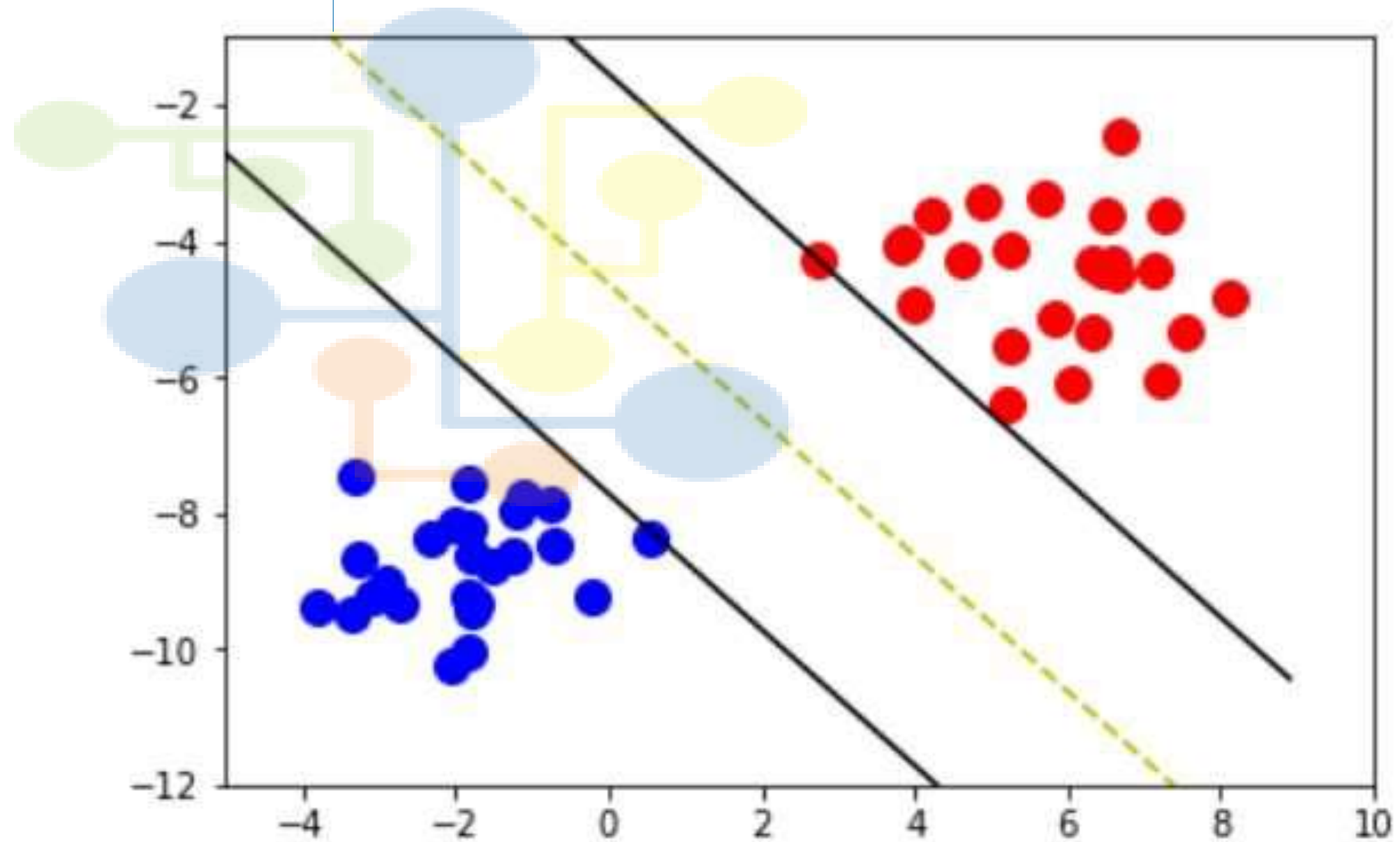


Como Funciona o Algoritmo SVM Para Classificação

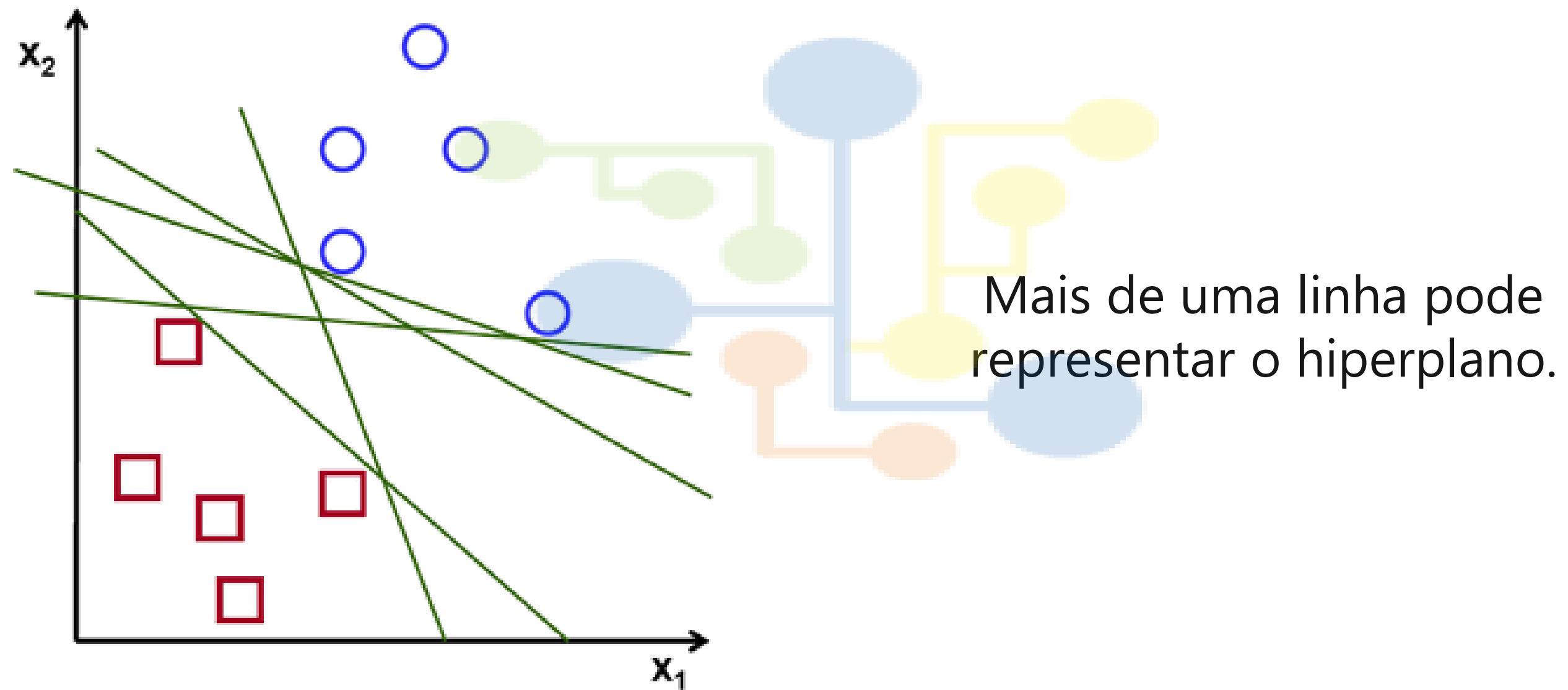


Hiperplano e Vetores de Suporte

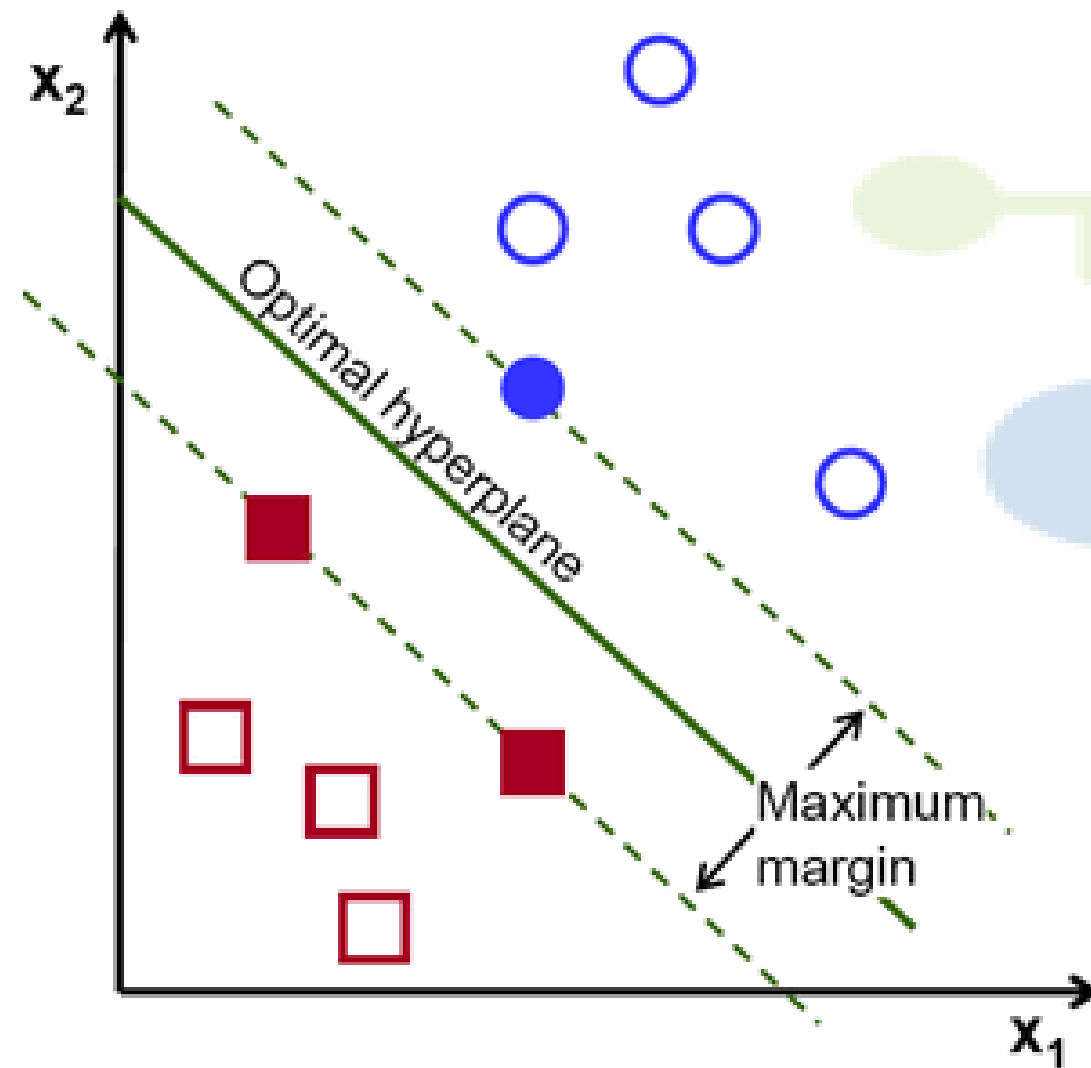
Hiperplano



Hiperplano e Vetores de Suporte



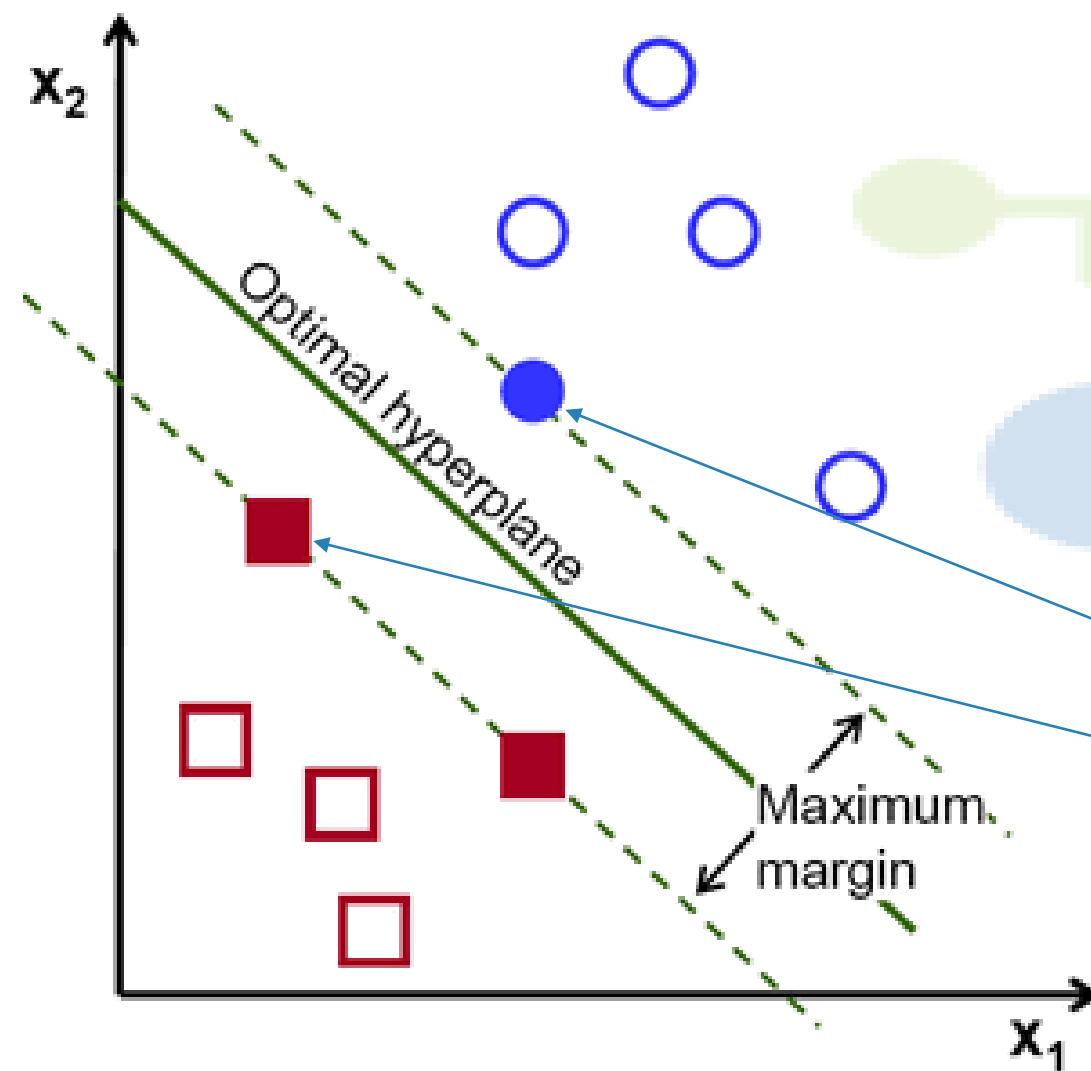
Hiperplano e Vetores de Suporte



O algoritmo seleciona a linha, ou seja o hiperplano, que maximiza as margens entre as classes.



Hiperplano e Vetores de Suporte



Os pontos que estão sobre as margens são conhecidos como vetores de suporte e é exatamente daí que vem o nome do algoritmo.

Vetores de Suporte



Matemática para Machine Learning

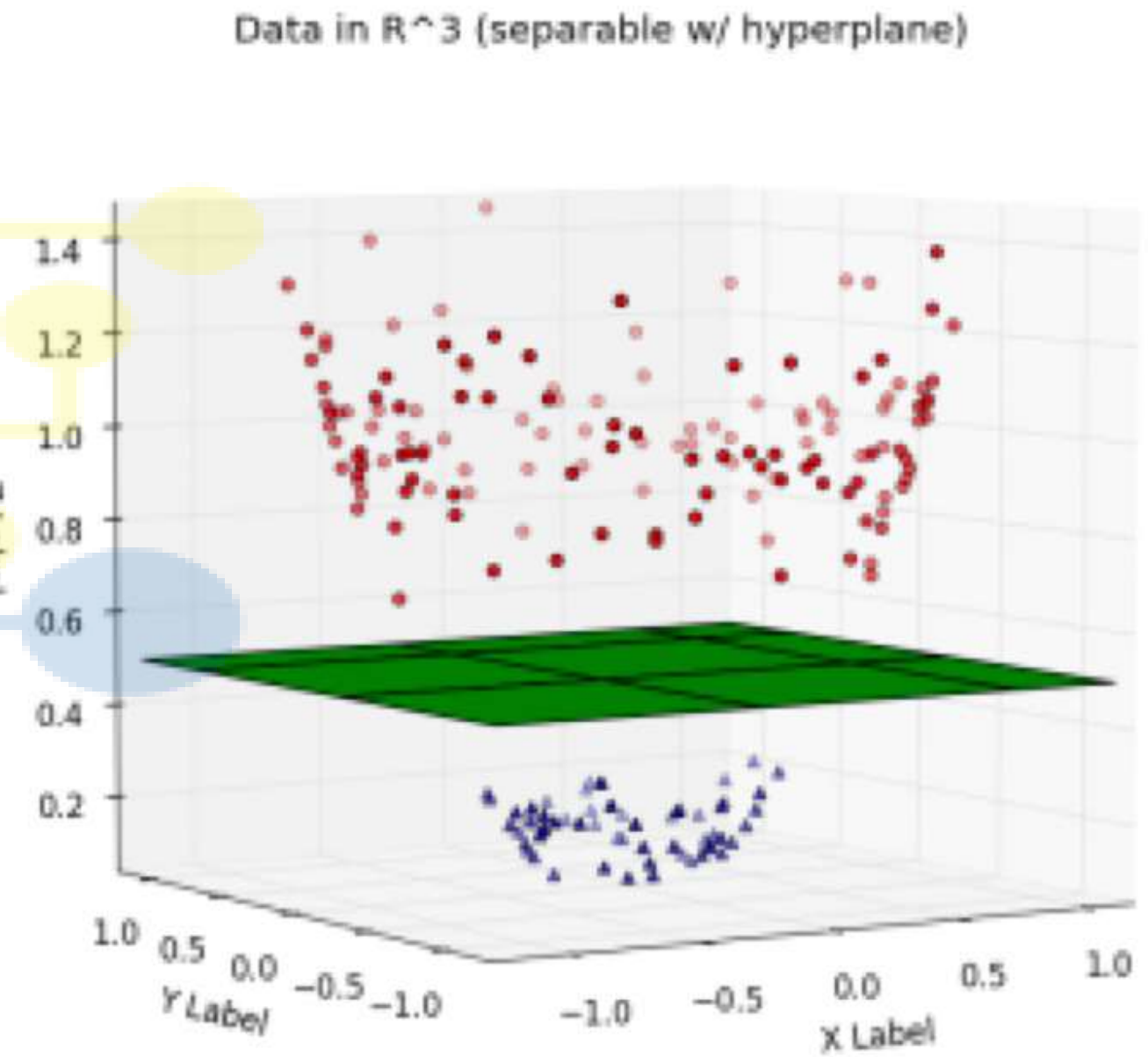
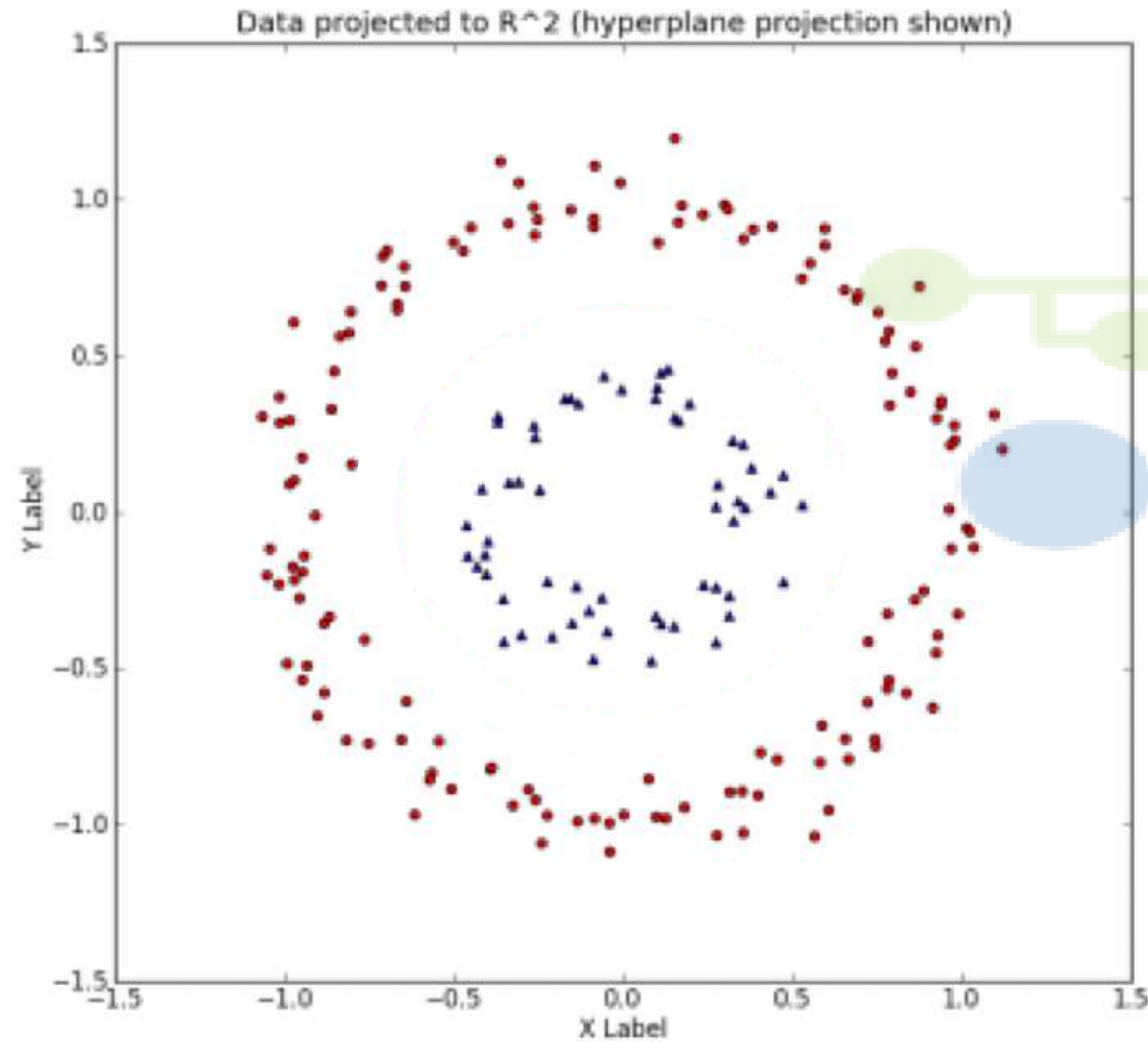


**E Quando os Dados Não São Linearmente
Separáveis?**



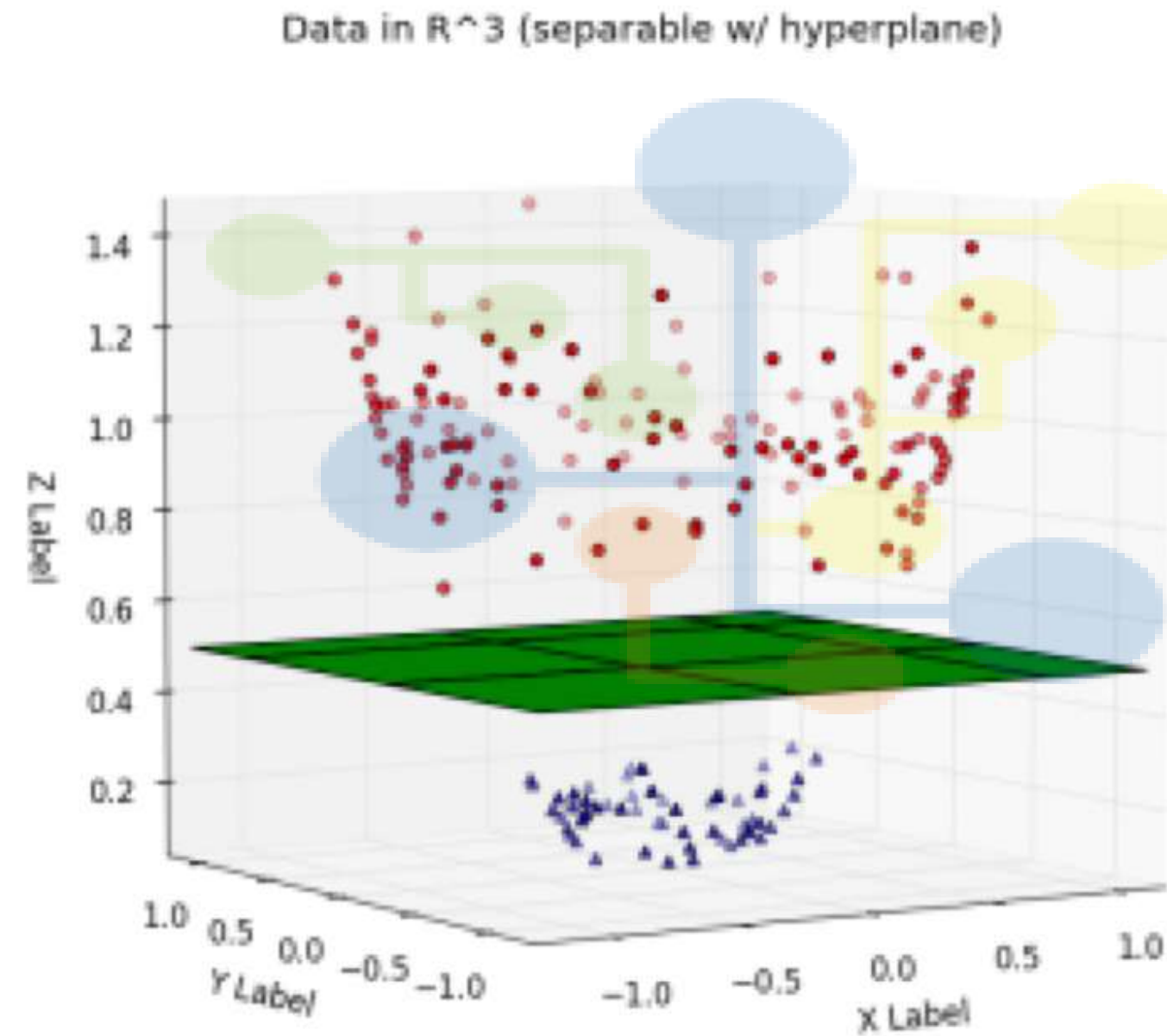


E Quando os Dados Não São Linearmente Separáveis?



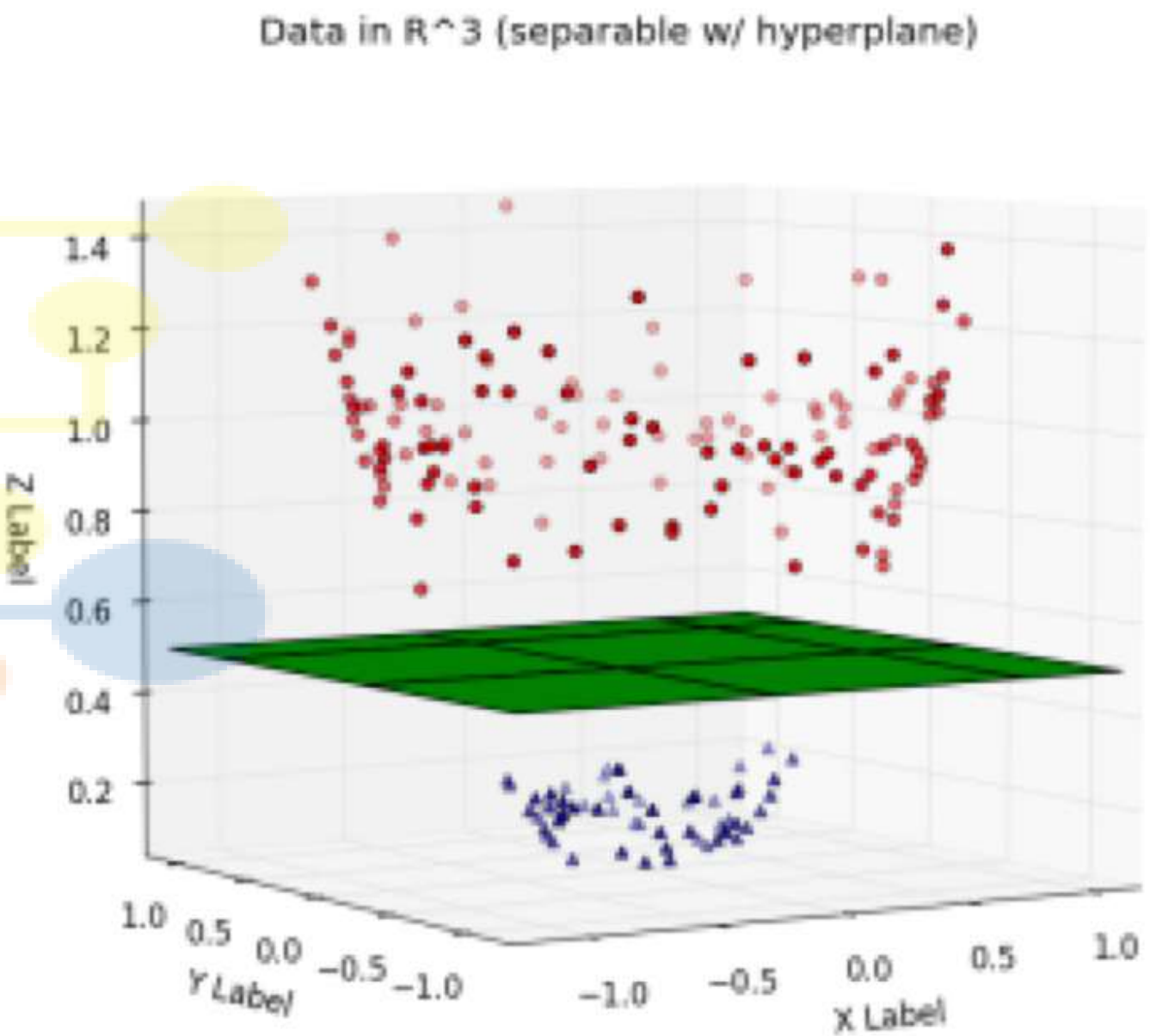
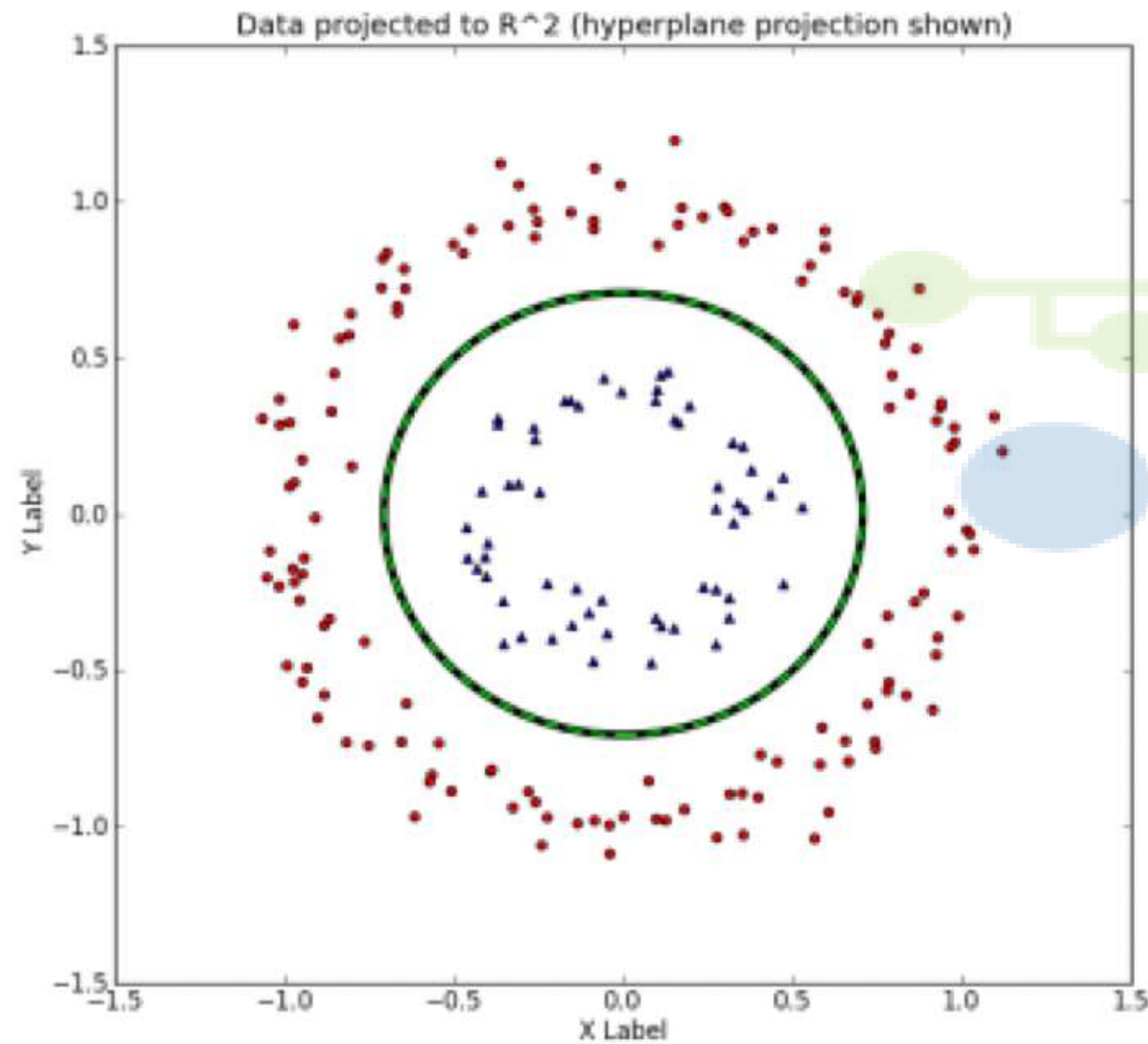


E Quando os Dados Não São Linearmente Separáveis?





E Quando os Dados Não São Linearmente Separáveis?



É um prazer ter você aqui!

Muito Obrigado!

Pela Confiança em Nosso Trabalho.

Continue Trilhando Uma Excelente Jornada de Aprendizagem!



Data Science Academy