VINI CARIDÁ

Human-Centered Data & AI

Vinicius Caridá, Ph.D.

Data Science Manager, Itaú Unibanco

MBA Professor, FIAP

GDE – Machine Learning

@vinicius caridá

@vfcarida

@vinicius caridá

@vfcarida

@vinicius caridá

@vfcarida

VINI CARIDÁ

Human-Centered Data & AI

SVM

Support Vector Machine

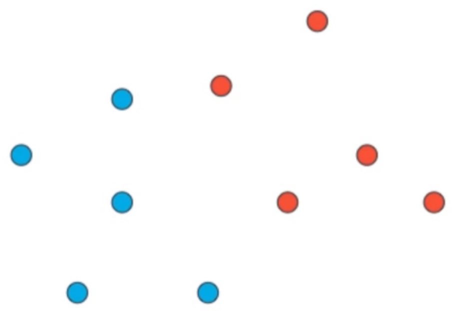
SVM | Support Vector Machine

- Proposto em 79 por Vladimir Vapnik
- Um dos mais importantes acontecimentos na área de reconhecimento de padrões nos últimos 15 anos.
- Tem sido largamente utilizado com sucesso para resolver diferentes problemas.

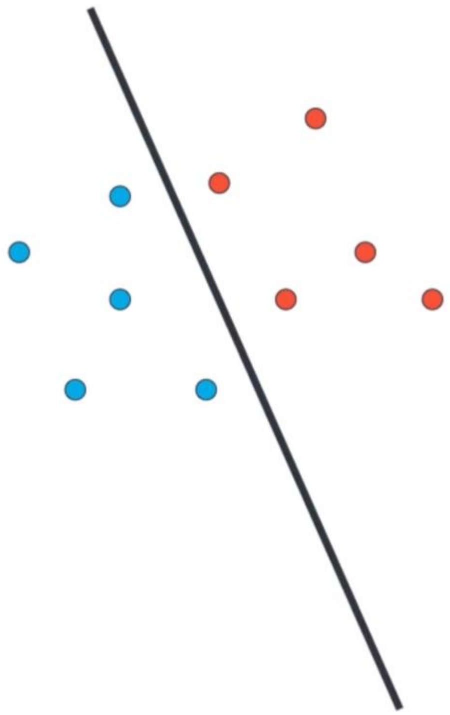
SVM | Support Vector Machine

- É um algoritmo de aprendizado de máquina supervisionado que é utilizado, principalmente, para problemas de classificação mas também pode ser usado para regressão (SVR); https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
- Com a existência de duas ou mais classes rotuladas de dados, ele age como um classificador discriminativo, formalmente definido por um hiperplano ideal que separa todas as classes;
- Novos exemplos que são então mapeados para o mesmo espaço podem ser categorizados com base em qual lado da lacuna eles se encontram.

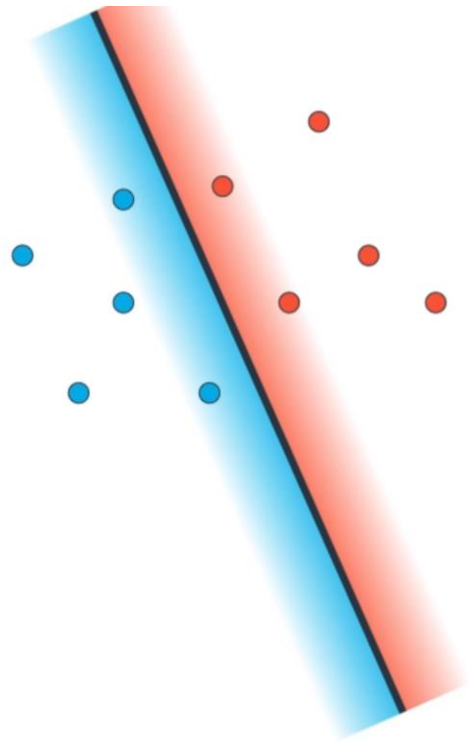
SVM | Support Vector Machine



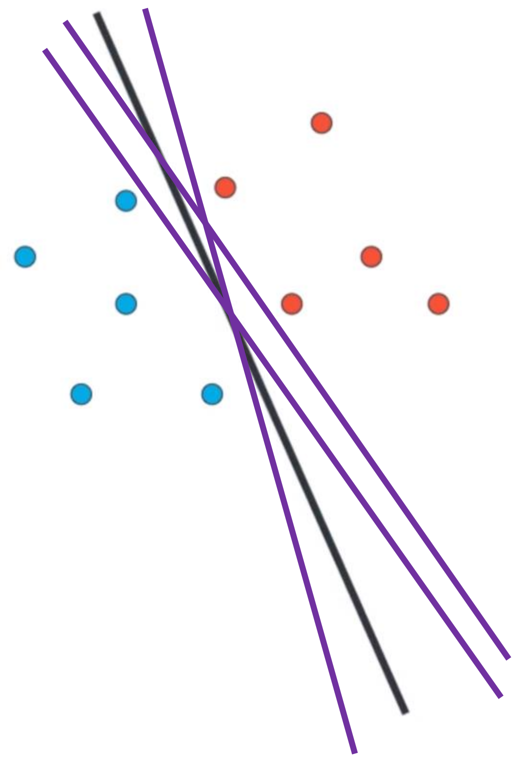
SVM | Support Vector Machine



SVM | Support Vector Machine



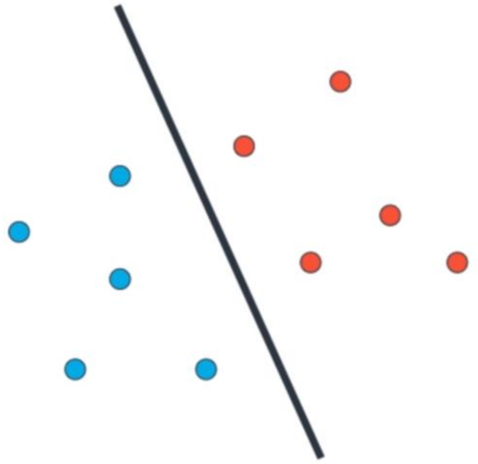
SVM | Support Vector Machine

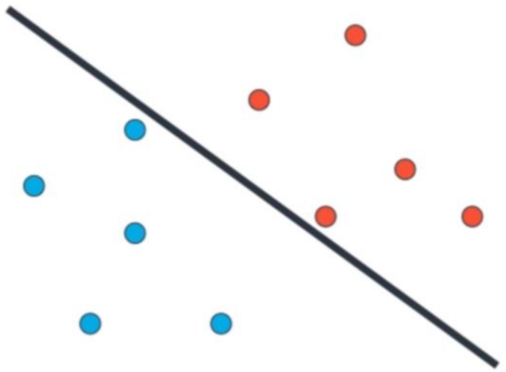


SVM | Support Vector Machine

SVM | Support Vector Machine

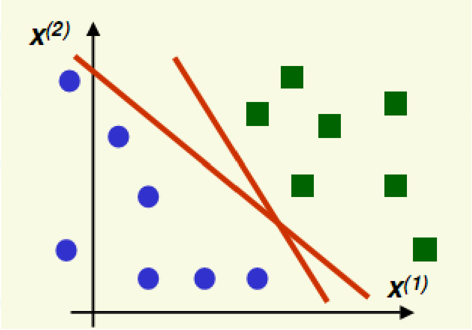
Qual linha é melhor?





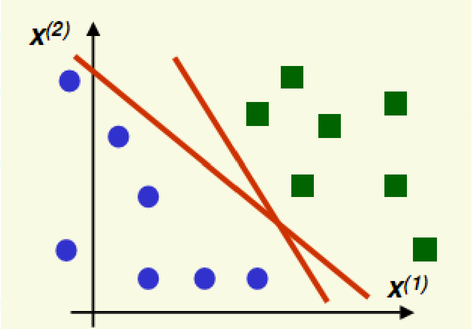
SVM | Support Vector Machine

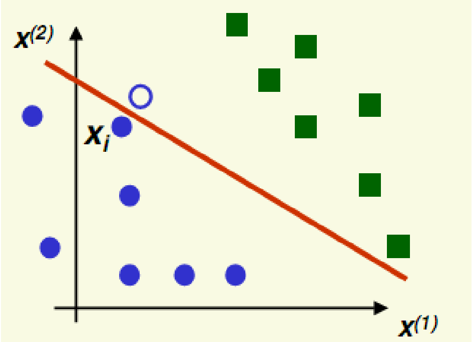
Qual a fronteira que deve ser escolhida??



SVM | Support Vector Machine

Qual a fronteira que deve ser escolhida??



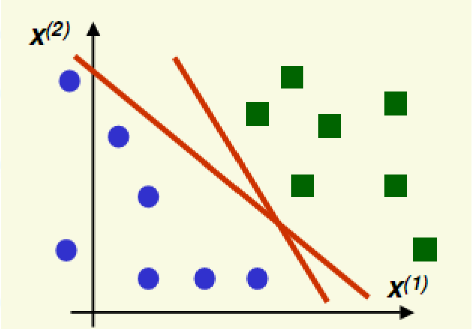


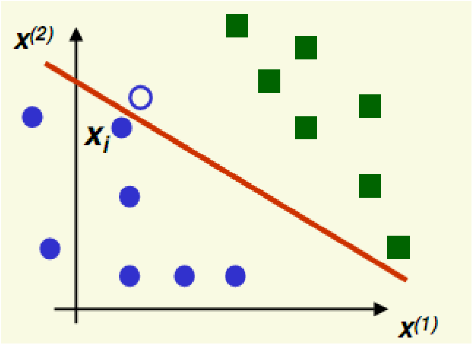
Como a linha está bem próxima da classe azul, seu **poder de generalização é baixo**

Note que um novo elemento (dados não usados no treinamento), bem próximo de um azul será classificado erroneamente.

SVM | Support Vector Machine

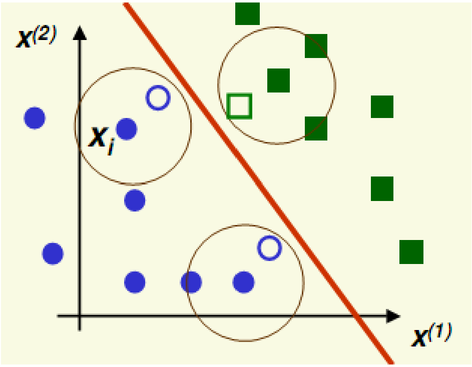
Qual a fronteira que deve ser escolhida??





Como a linha está bem próxima da classe azul, seu **poder de generalização é baixo**

Note que um novo elemento (dados não usados no treinamento), bem próximo de um azul será classificado erroneamente.

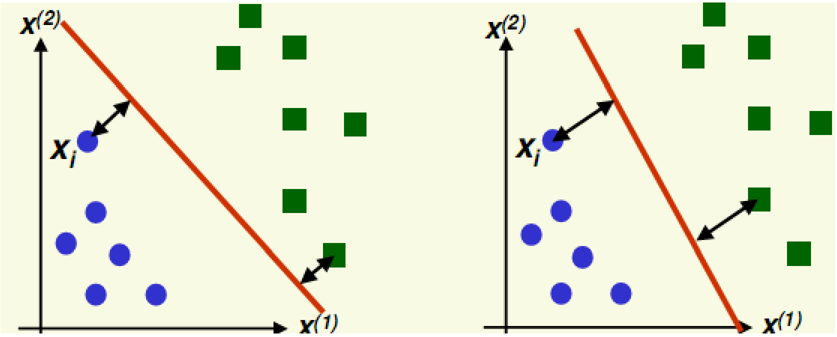


Podemos notar que o poder de generalização é bem melhor.

Novos dados são corretamente classificados, pois temos uma fronteira mais distante dos dados de treinamento.

SVM | Support Vector Machine

- O conceito por traz do SVM é a maximização da margem, ou seja, maximizar a distância da margem dos dados de treinamento

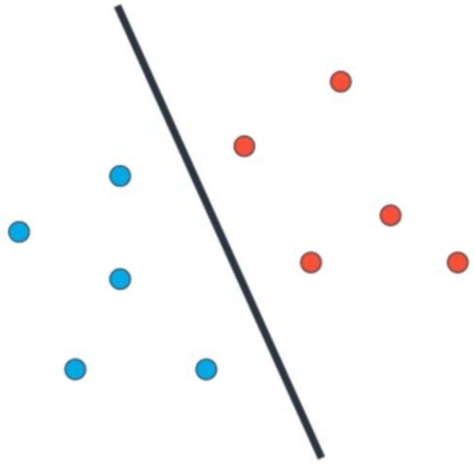


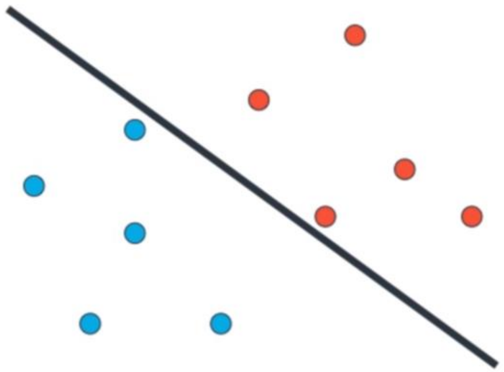
Distância Pequena

Distância Grande

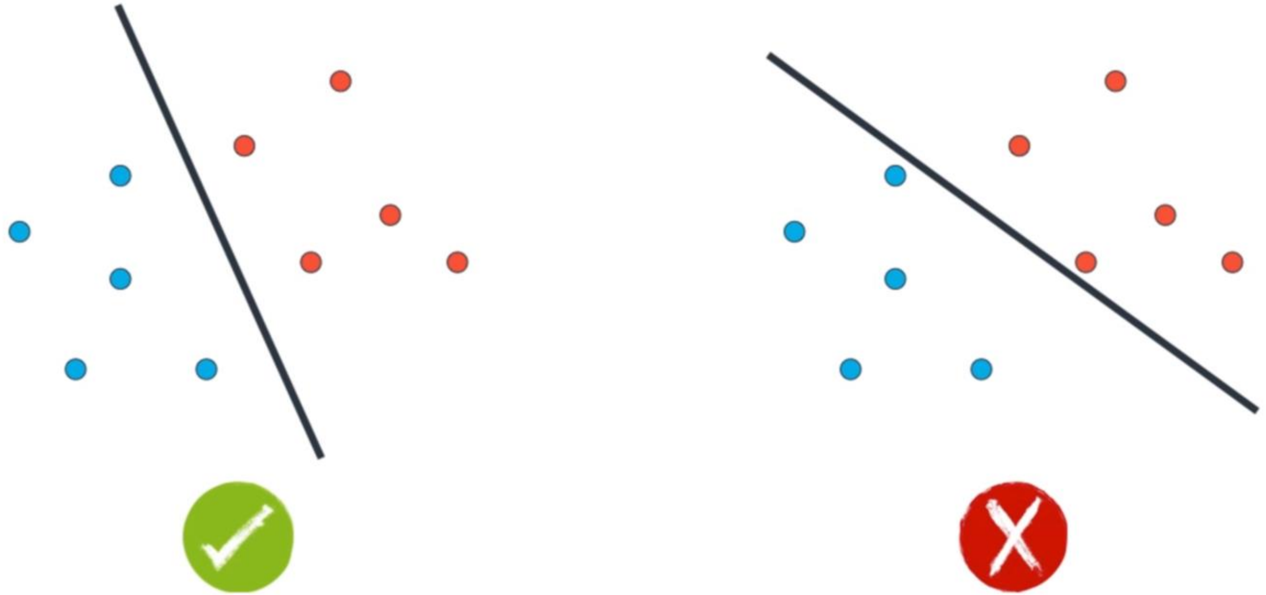
SVM | Support Vector Machine

Qual linha é melhor?



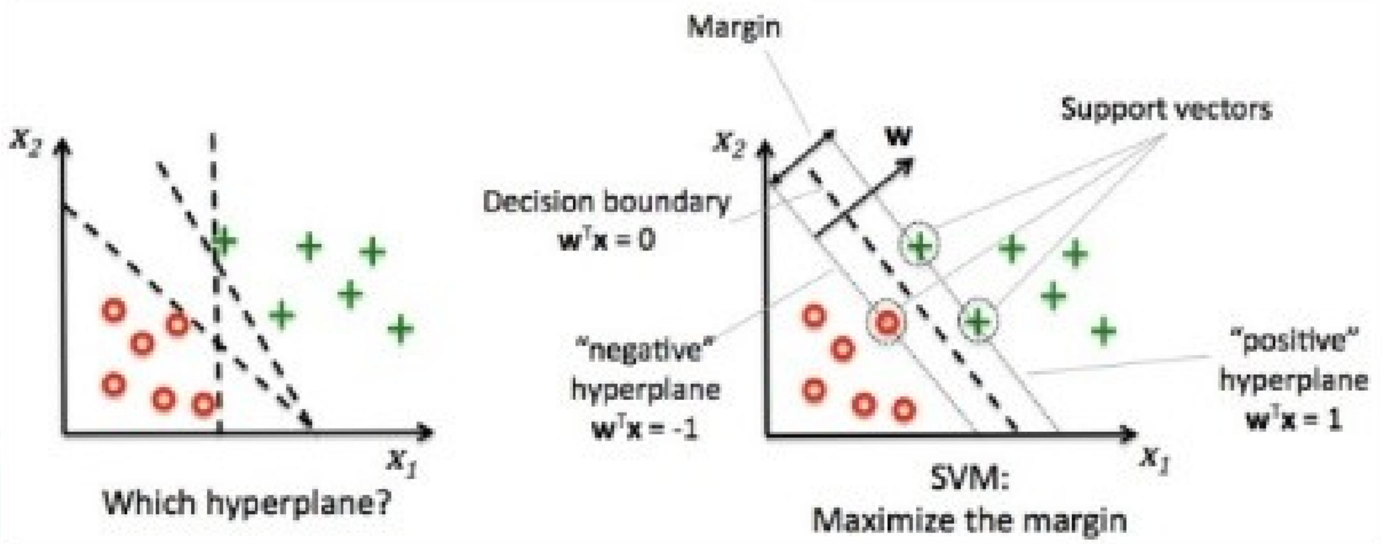


SVM | Support Vector Machine



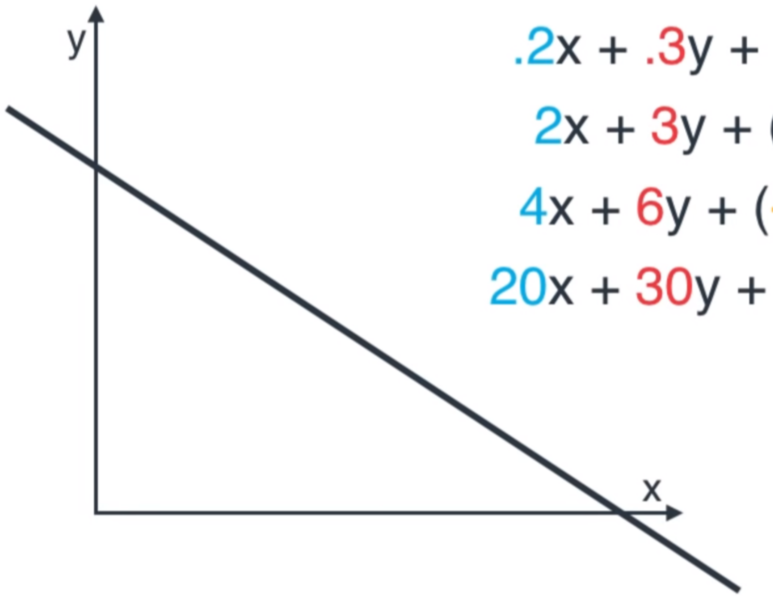
SVM | Support Vector Machine

- Os vetores de suporte são os pontos de dados mais próximos do hiperplano; os pontos de um conjunto de dados que, se removidos, alterariam a posição do hiperplano divisor;
- Por causa disso, eles podem ser considerados os elementos críticos de um conjunto de dados, são eles que nos ajudam a construir nosso SVM.



SVM | Support Vector Machine

Como separar as linhas?



$$.2x + .3y + (-.6) = 0$$

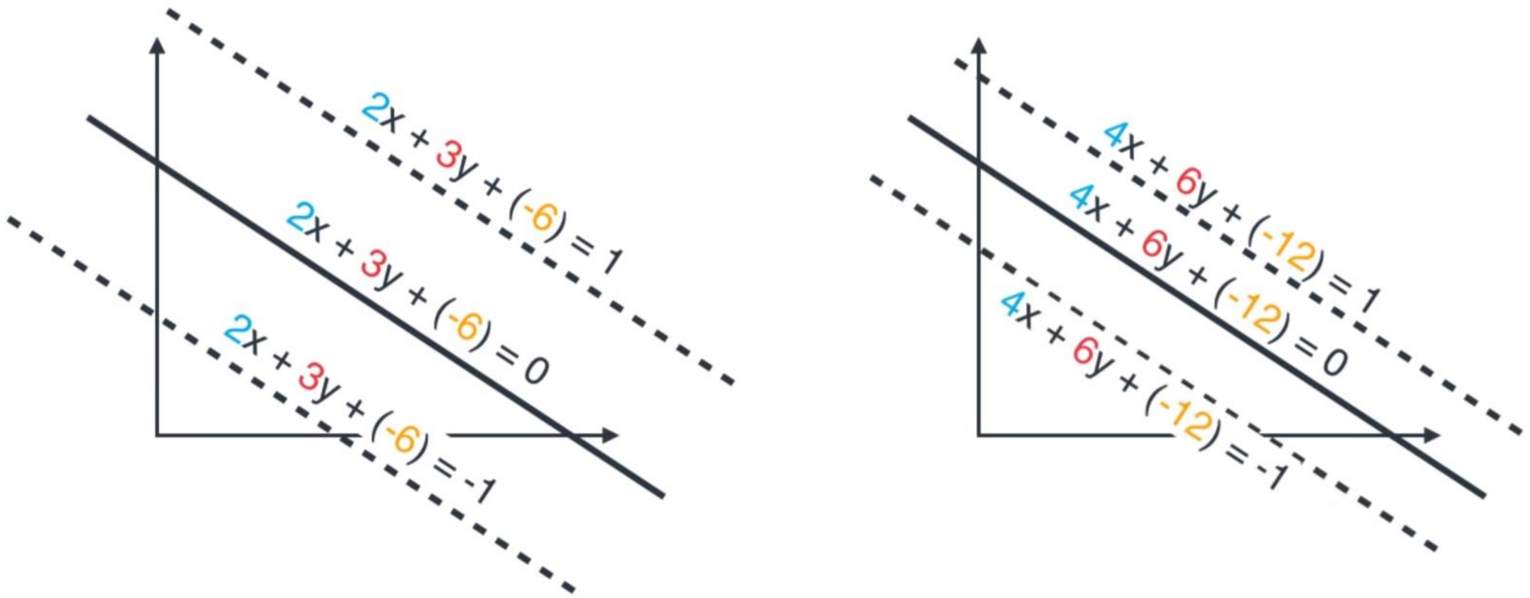
$$2x + 3y + (-6) = 0$$

$$4x + 6y + (-12) = 0$$

$$20x + 30y + (-60) = 0$$

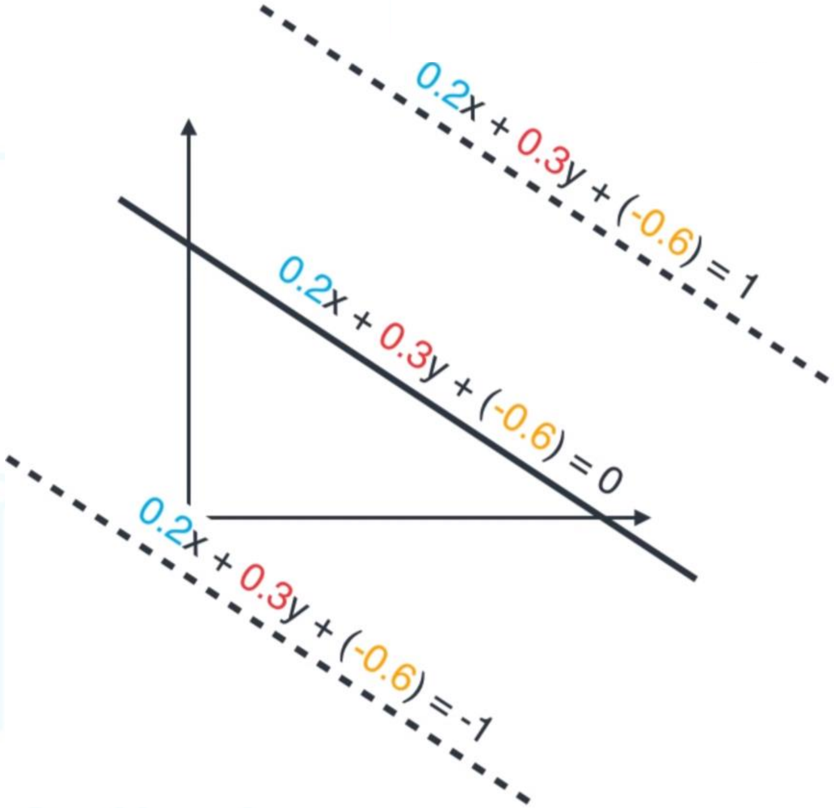
SVM | Support Vector Machine

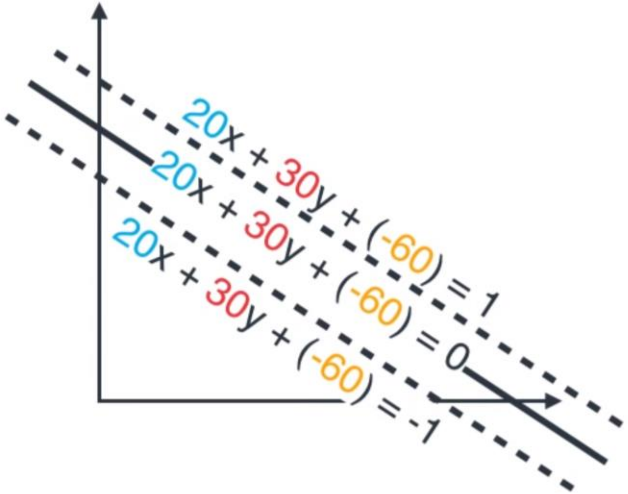
Como separar as linhas?



SVM | Support Vector Machine

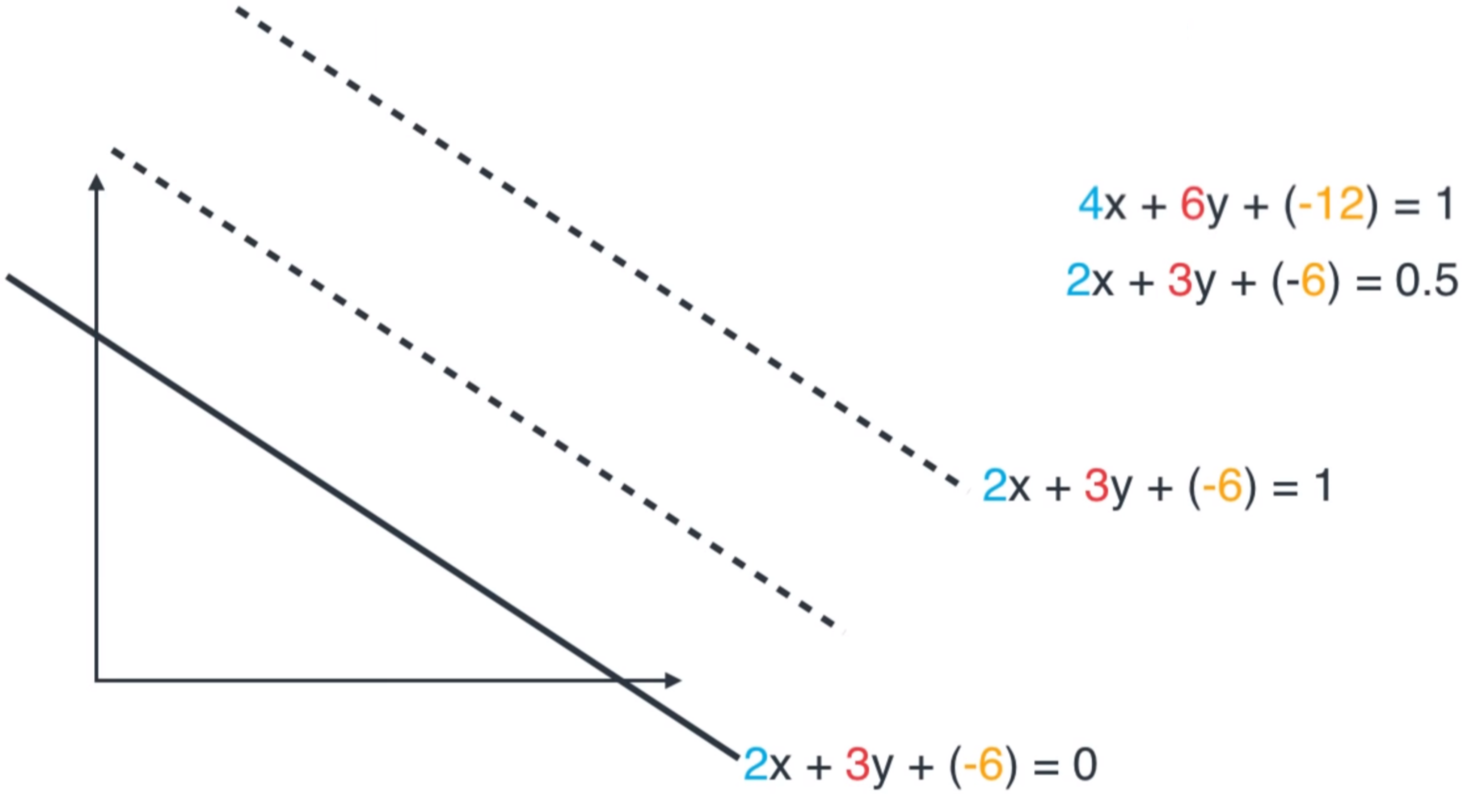
Como separar as linhas?





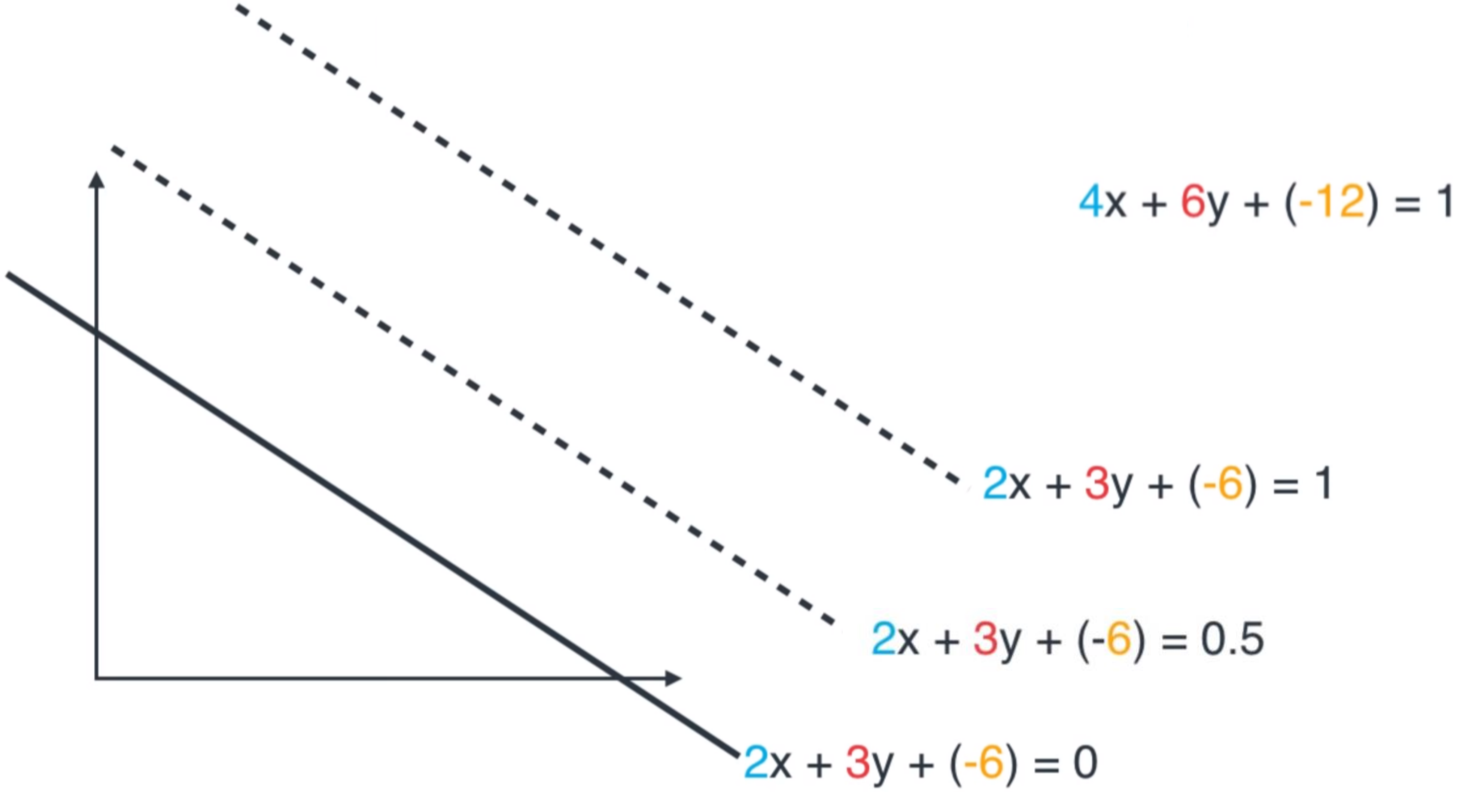
SVM | Support Vector Machine

Como separar as linhas?



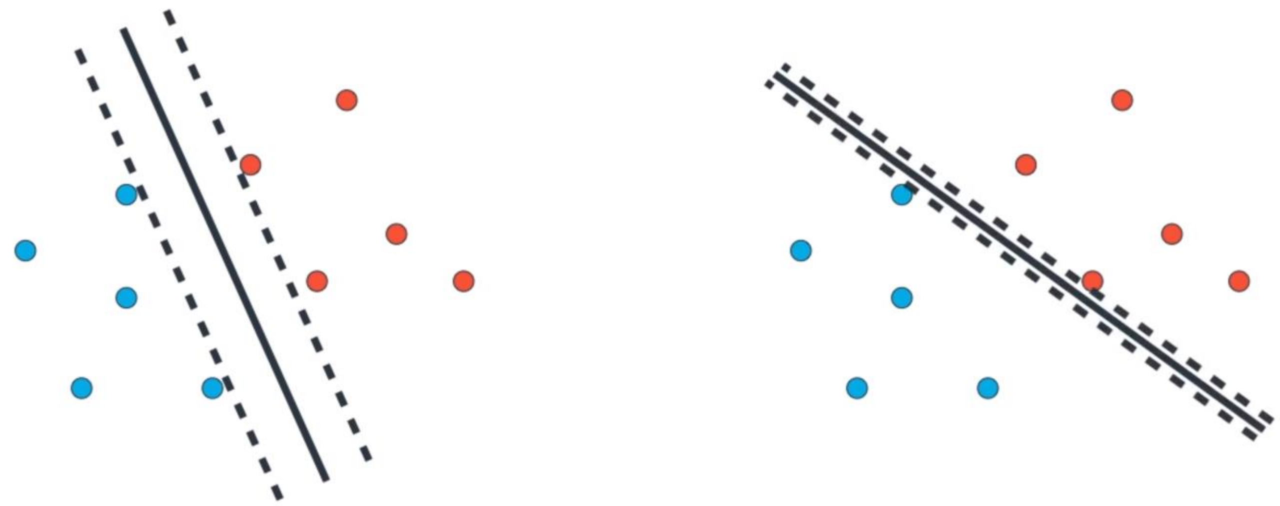
SVM | Support Vector Machine

Como separar as linhas?



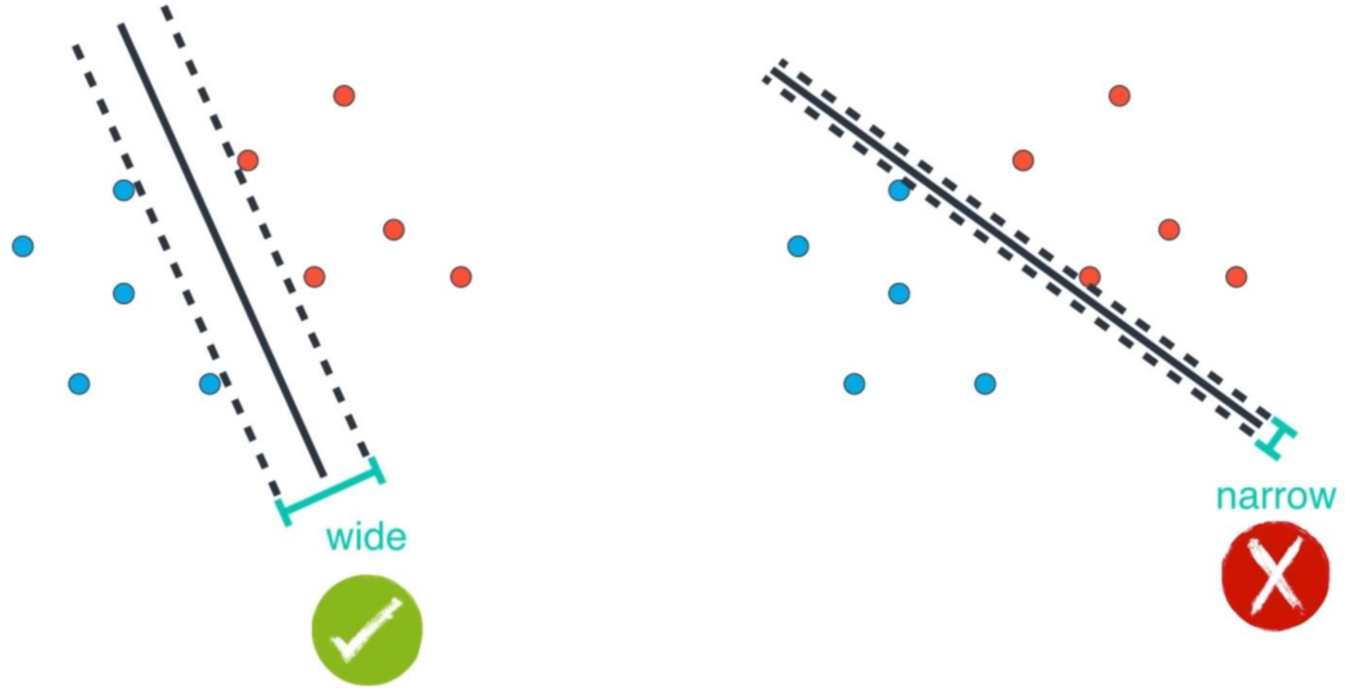
SVM | Support Vector Machine

Qual linha é melhor?

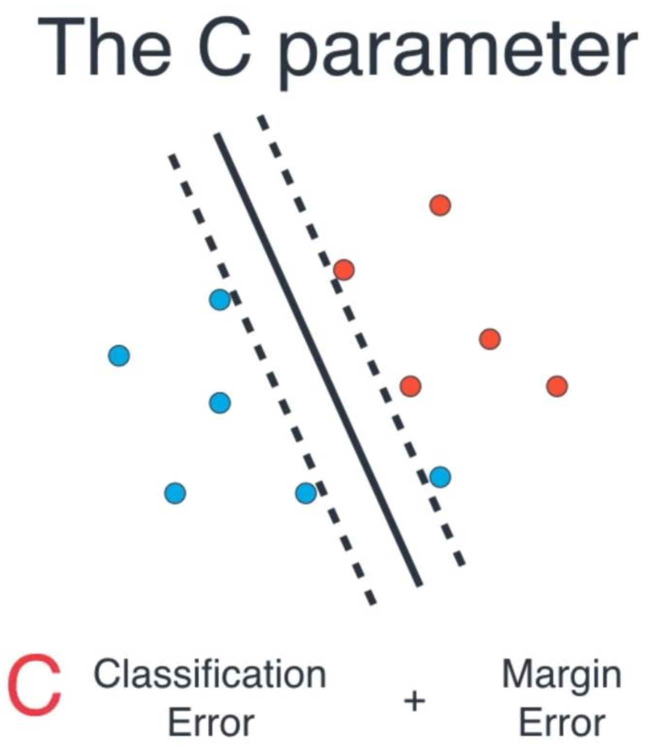


SVM | Support Vector Machine

Qual linha é melhor?

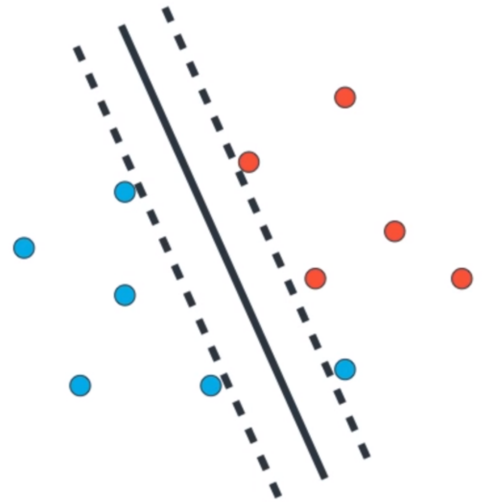


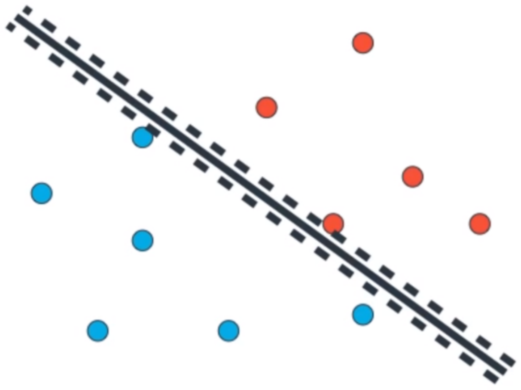
SVM | Support Vector Machine

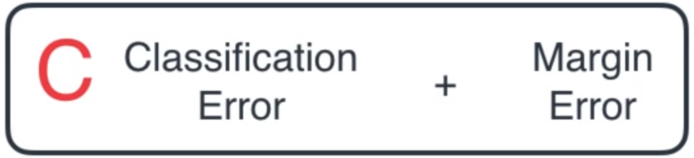


SVM | Support Vector Machine

Qual linha é melhor?

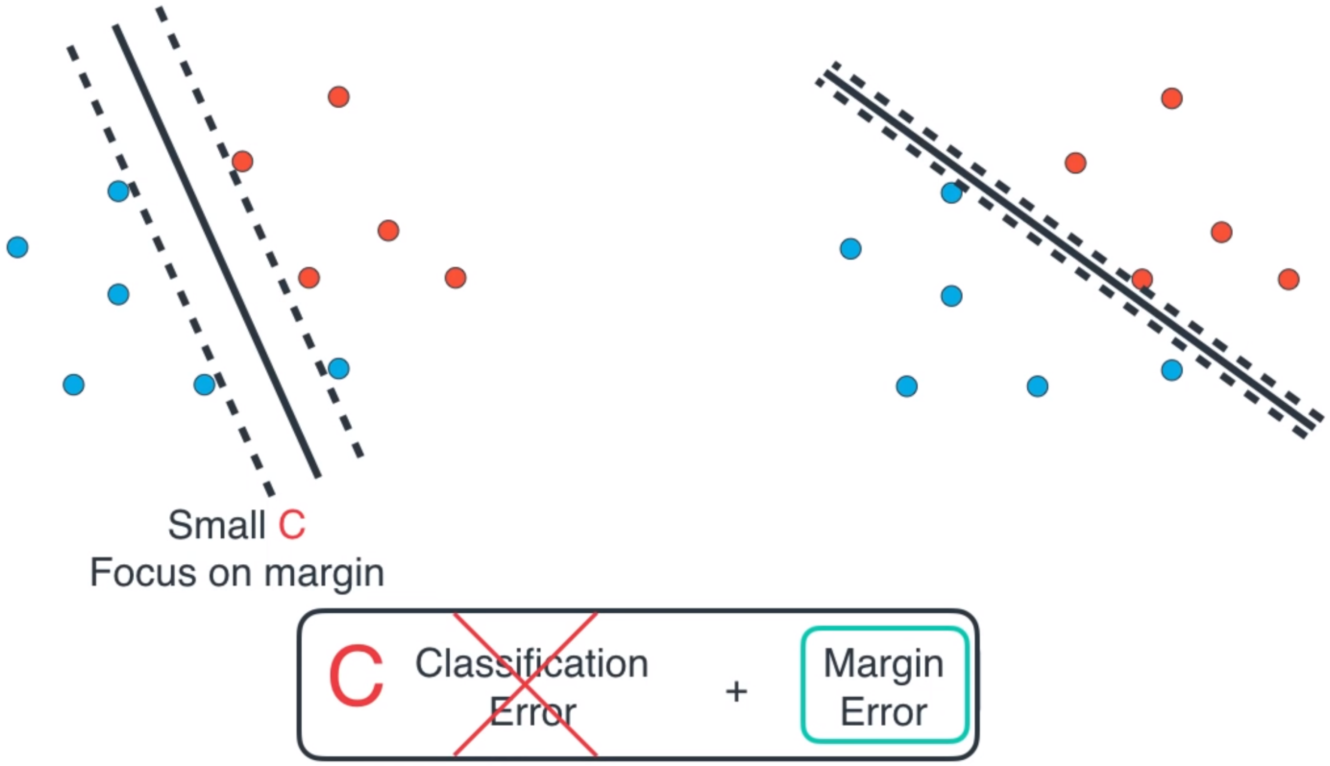






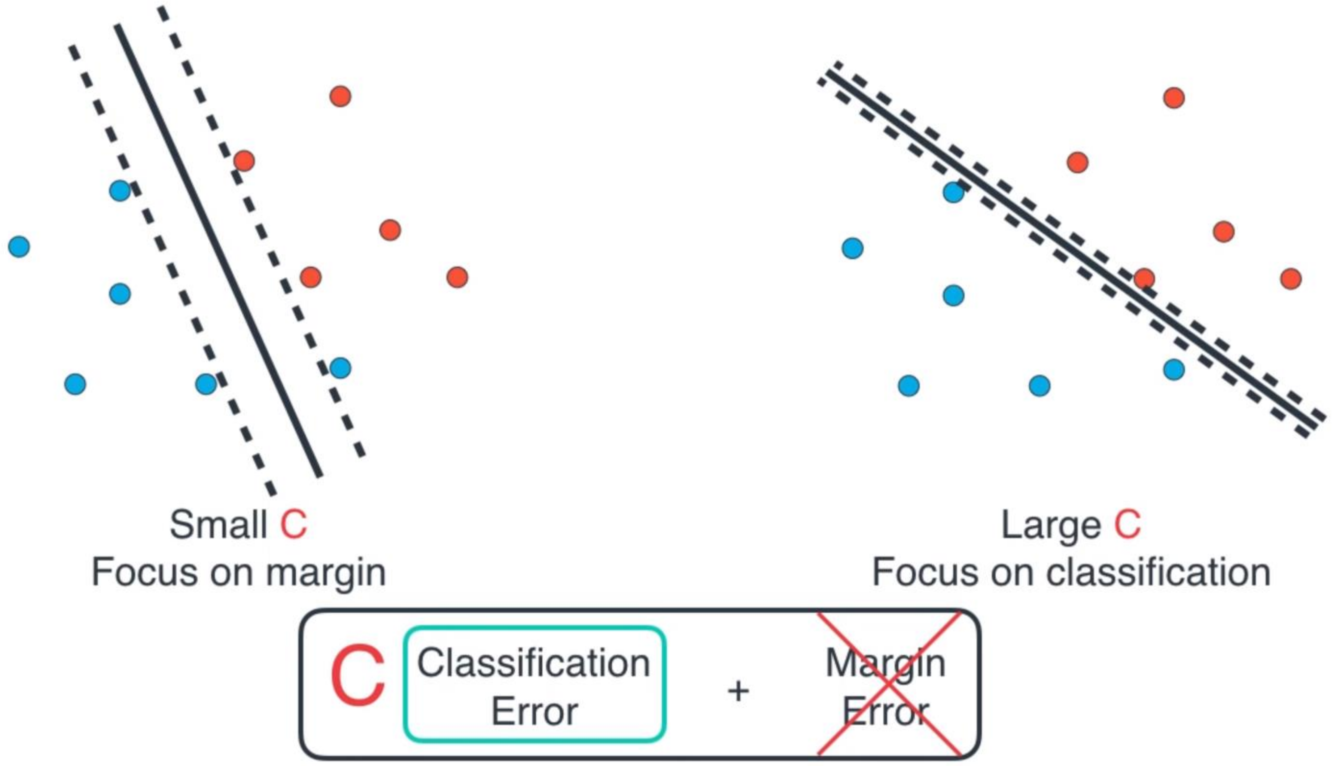
SVM | Support Vector Machine

Qual linha é melhor?

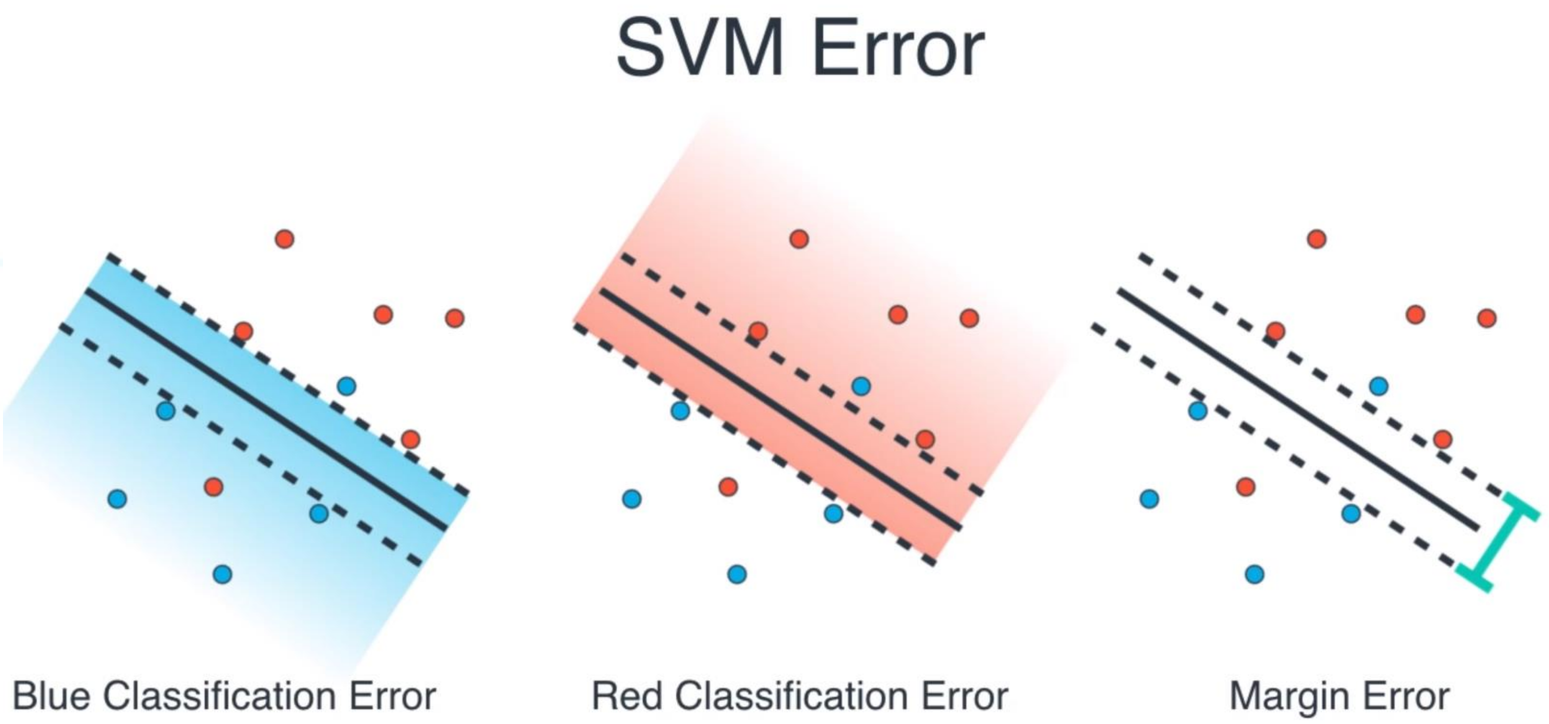


SVM | Support Vector Machine

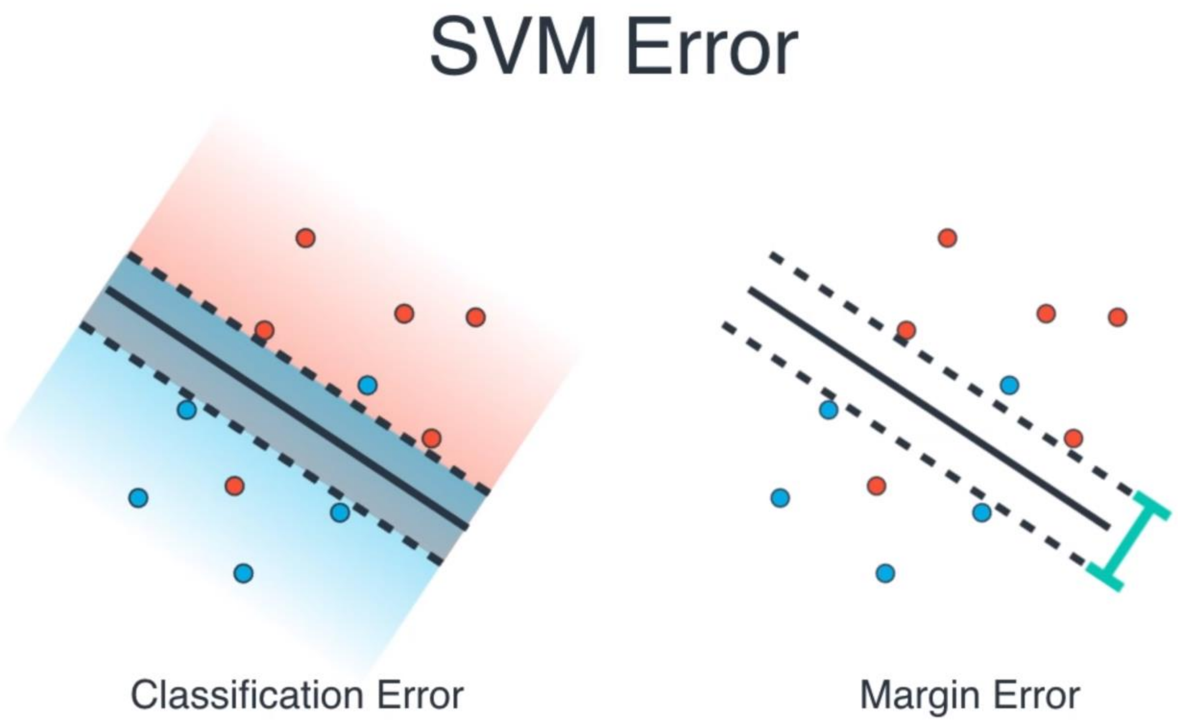
Qual linha é melhor?



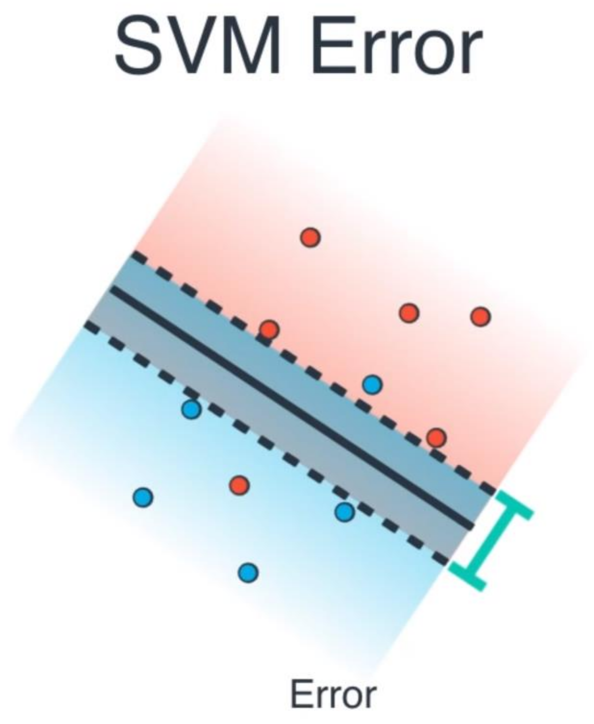
SVM | Support Vector Machine



SVM | Support Vector Machine



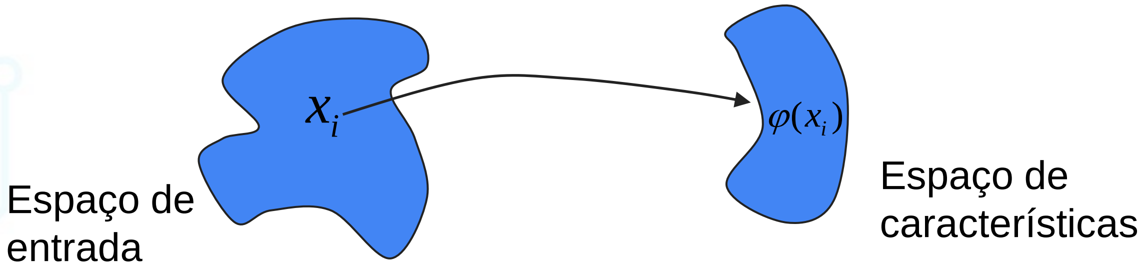
SVM | Support Vector Machine



SVM | Support Vector Machine

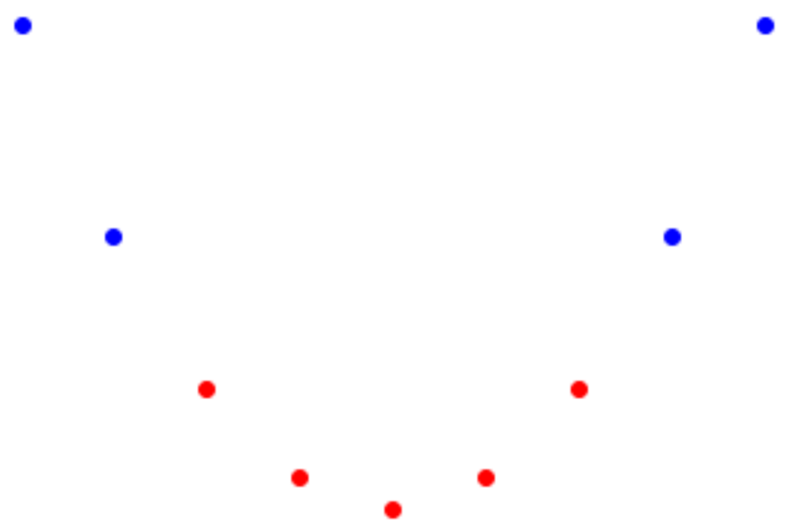
- A grande maioria dos problemas reais não são linearmente separáveis.
- A pergunta então é: “Como resolver problemas que não são linearmente separáveis com um classificador linear?”

Projetar os dados em um espaço onde os dados são linearmente separáveis.

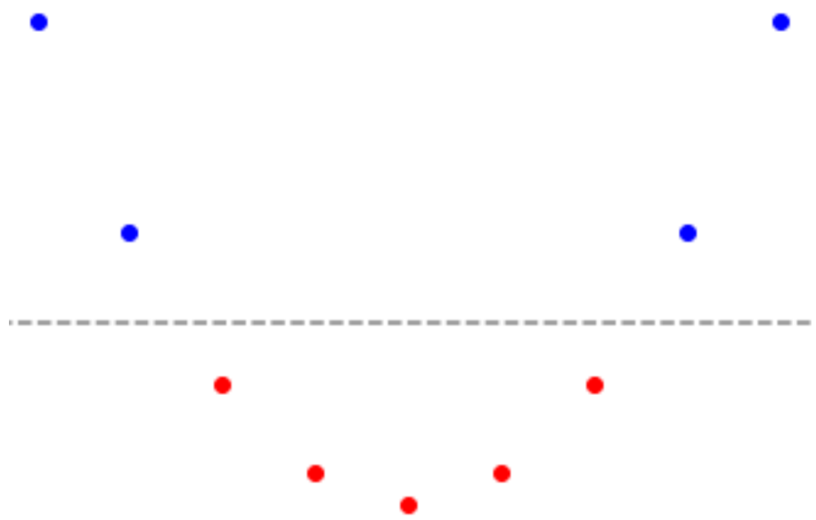


SVM | Support Vector Machine

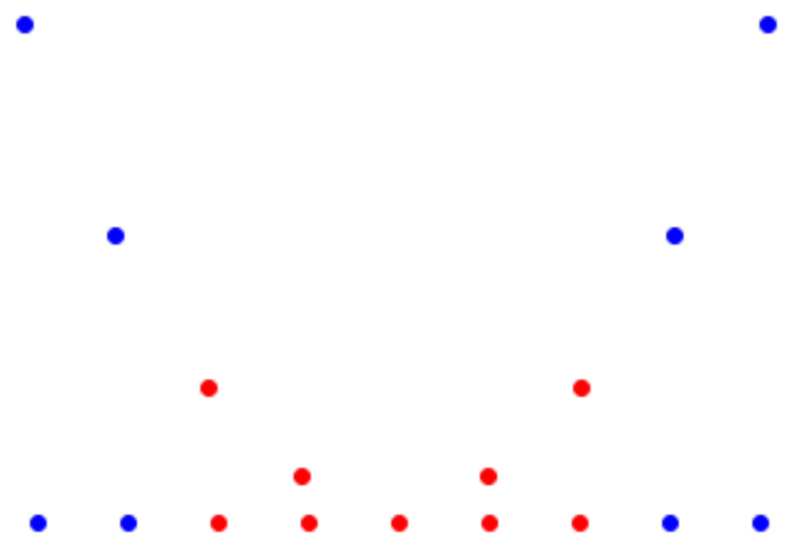
SVM | Support Vector Machine



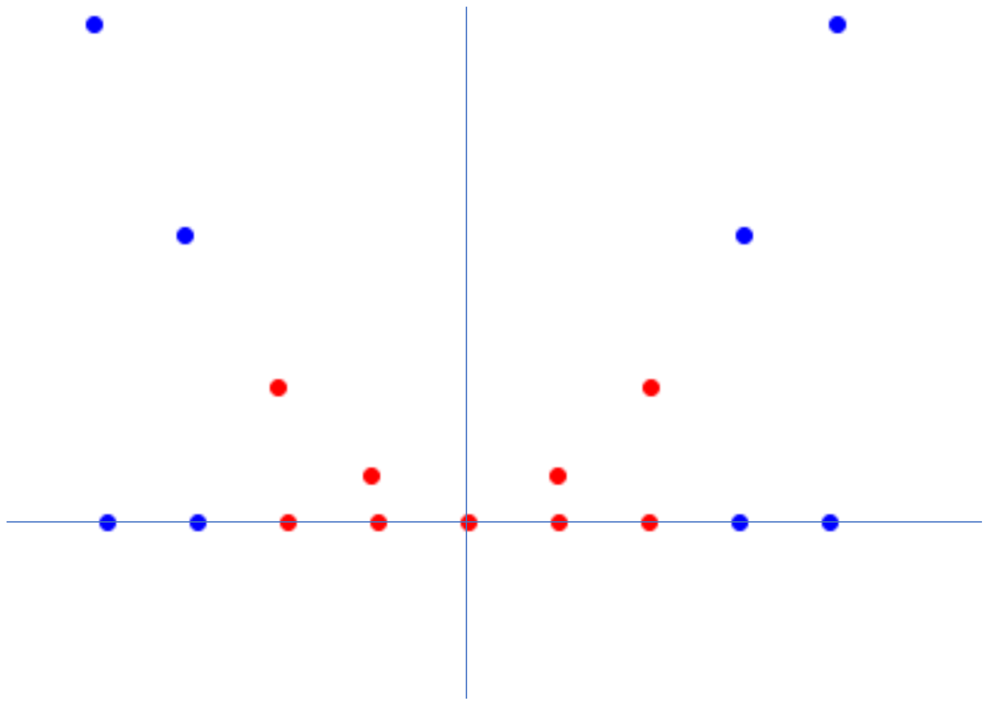
SVM | Support Vector Machine



SVM | Support Vector Machine



SVM | Support Vector Machine



$$x = [x, x^2]$$

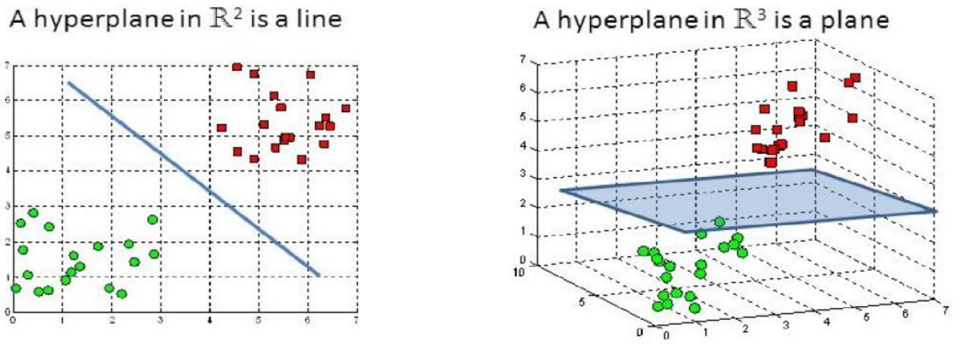
$$1 = [1,1]$$

$$-1 = [-1,1]$$

SVM | Support Vector Machine

- A geometria nos diz que um hiperplano é um subespaço de uma dimensão menor que seu espaço de ambiente; Por exemplo, um hiperplano de um espaço n-dimensional é um subconjunto plano com dimensão n-1;

- Simplificando, é como se um hiperplano separasse o espaço em dois meios espaços (classificador binário).



A hyperplane in $\mathbb{R}^n$ is an n-1 dimensional subspace

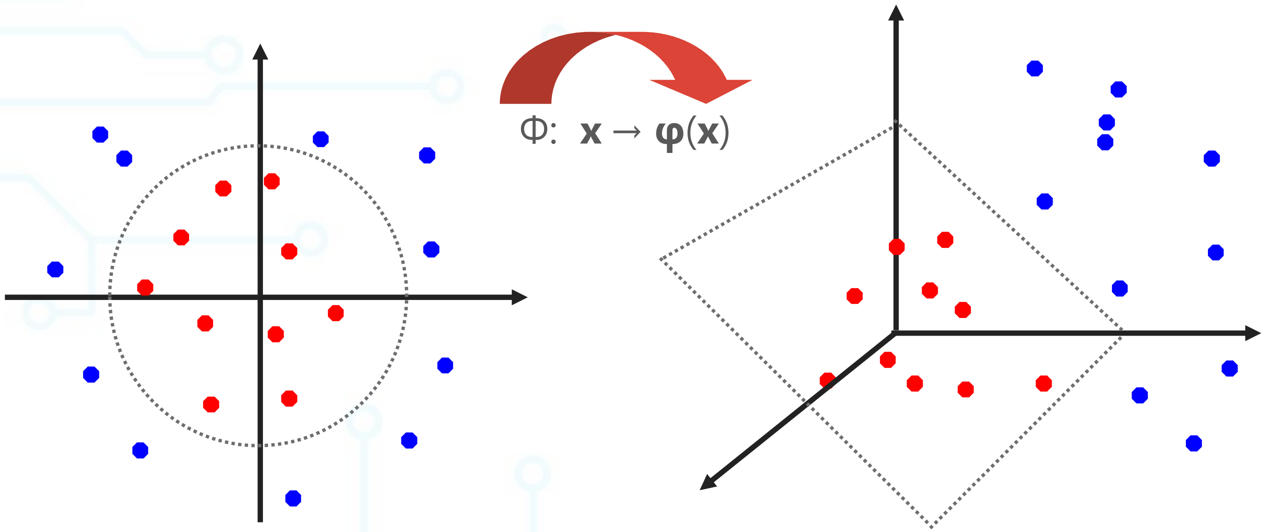
SVM | Support Vector Machine

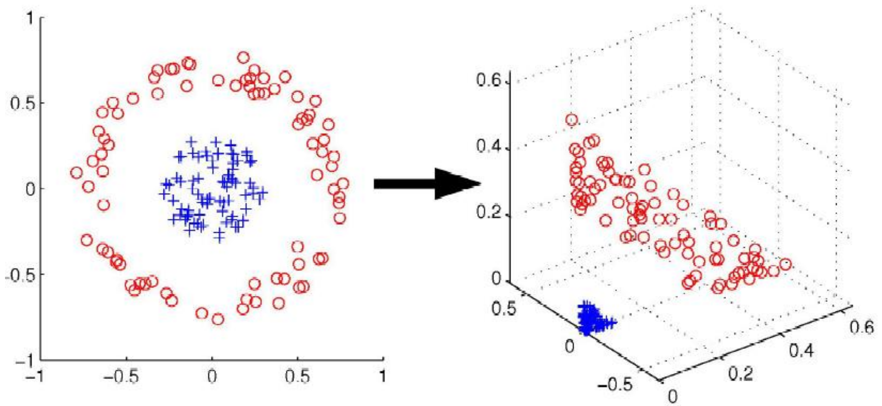
$$f(x) = \sum_i \alpha_i y_i K(x, x_i) + b$$

- A função de decisão pode ser descrita pela fórmula acima, na qual:
  - K é a função de kernel,
  - α e b são os parâmetros encontrados durante o treinamento,
  - x_i e y_i são os vetores de características e o *label* da classe respectivamente.

SVM | Support Vector Machine

$$
\begin{array}{llll}
\phi: & \Re^2 & \longrightarrow & \Re^3 \\
& (x_1, x_2) & \longmapsto & (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)
\end{array}
$$





SVM

Anexos

SVM | Support Vector Machine

Considere que o conjunto de treinamento $\{(\mathbf{x}_i, y_i)\}_{i=1..N}$, $\mathbf{x}_i \in \mathbf{R}^d$, $y_i \in \{-1, 1\}$, seja separado por um hiperplano de margem ρ. Então para cada $(\mathbf{x}_i, y_i)$ temos:

$$\begin{array}{ll} \mathbf{w}^T\mathbf{x}_i + b \leq -\rho/2 & \text{se } y_i = -1 \\ \mathbf{w}^T\mathbf{x}_i + b \geq \rho/2 & \text{se } y_i = 1 \end{array} \quad \Leftrightarrow \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq \rho/2$$

Para cada $\mathbf{x}_s$ a inequação acima se torna uma equação. Após reescalar $\mathbf{w}$ e b por $\rho/2$ na equação, a distância entre cada $\mathbf{x}_s$ e o hiperplano é:

$$r = \frac{y_s(\boldsymbol{w}^T\boldsymbol{x}_s + b)}{\|\boldsymbol{w}\|} = \frac{1}{\|\boldsymbol{w}\|}$$

Então a margem pode ser expressa por meio de $\mathbf{w}$ e b como:

$$\rho = 2r = \frac{2}{\|\boldsymbol{w}\|}$$

SVM | Support Vector Machine

Então podemos formular o problema de otimização quadrática:

Encontrar **w** e b tal que

seja maximizado

e para todos os $(\mathbf{x}_i, y_i)$, $i=1..N$: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

Este pode ser reformulado como:

Encontrar **w** e b tal que

$\mathbf{\Phi}(\mathbf{w}) = ||\mathbf{w}||^2=\mathbf{w}^T\mathbf{w}$ seja minimizado

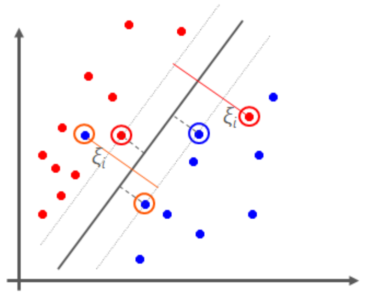
E para todo $(\mathbf{x}_i, y_i)$, $i=1..N$: $y_i\ (\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

SVM | Support Vector Machine

- Problema de otimização passa então a ser:

Encontrar $\mathbf{w}$ e b tal que
$\Phi(\mathbf{w}) = \mathbf{w}^T\mathbf{w} + C\Sigma\xi_i$ é minimizado
e para todo $(\mathbf{x}_i, y_i)$, $i=1..N$: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

- O parâmetro C é útil para controlar o superajuste (*overfitting*), balanceando a importância relativa de maximizar a margem e ajustá-la aos dados de treinamento.
- Escolher C via validação cruzada.



SVM | Support Vector Machine

Algumas opções de kernels:

Kernel	Inner Product Kernel
Linear	$K(x, y) = (x \cdot y)$
Gaussian	$K(x, y) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$
Polynomial	$K(x, y) = (x \cdot y)^p$
Tangent Hyperbolic	$K(x, y) = \tanh(x \cdot y - \Theta)$

- Escolher aquele que se ajusta melhor aos dados: fronteiras de decisão podem requerer kernels complexos;

- Maximizar a margem → pesos → vetores

Thanks !

Vinicius Fernandes Caridá

vfcarida@gmail.com

@vinicius caridá @vfcarida @vinicius caridá @vfcarida @vinicius caridá @vfcarida