"Para o prazer e para ser feliz, é que é preciso a gente saber tudo, formar alma, na consciência; para penar, não se carece."

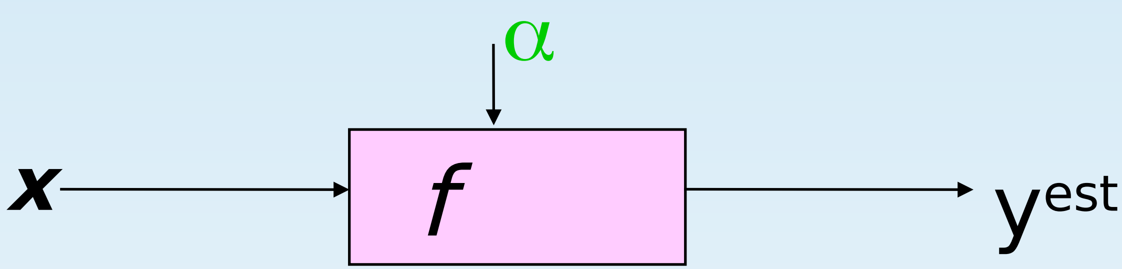
(Guimarães Rosa *in* **Grande Sertão: Veredas**,1956)

Roteiro da aula:

- **Máquinas de Vetores de Suporte (SVM)**

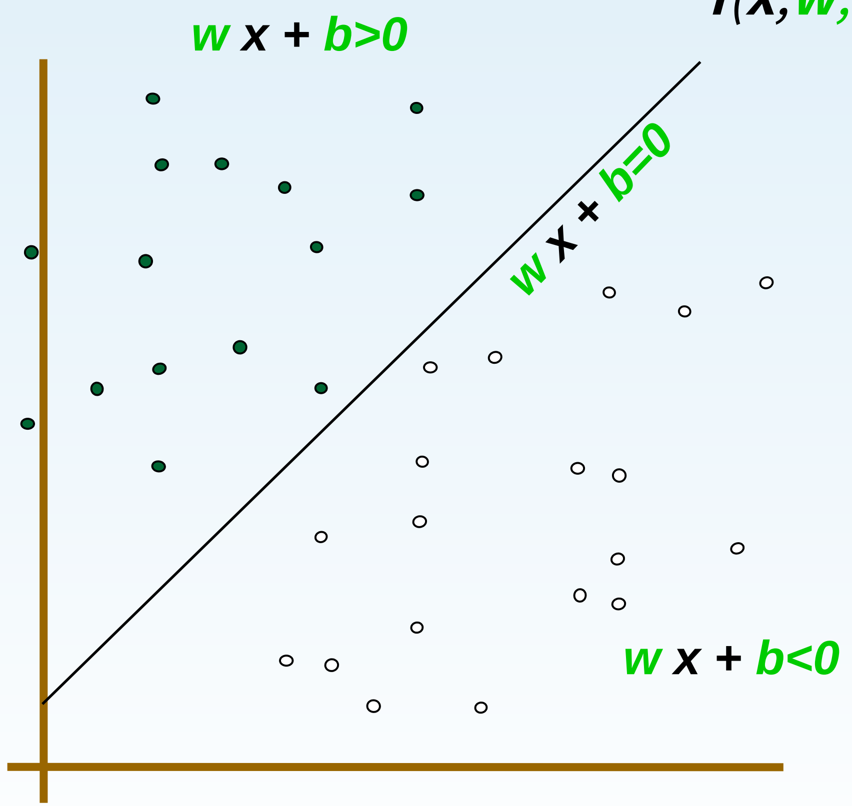
Máquinas de Vetores de Suporte. Material baseado, com ilustrações, no Cap. 9 de "Introduction to Statistical Learning", James, Witten, Hastie & Tibshirani, Springer, 2017.

De classificadores lineares por funções discriminantes



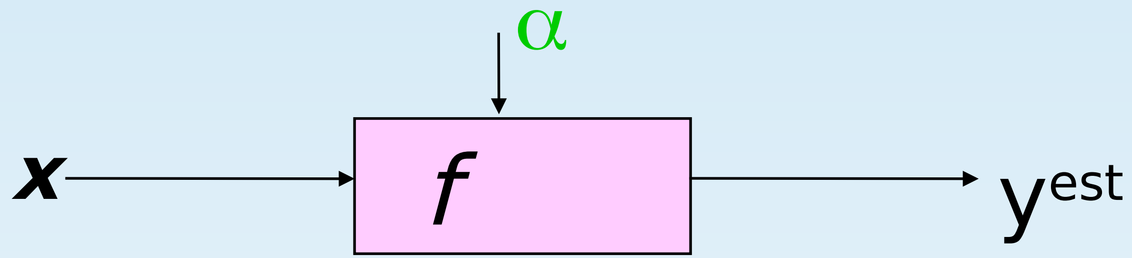
$$f(\mathbf{x}, w, b) = sign(w\,\mathbf{x} + b)$$

- ● indica +1
- ○ indica -1



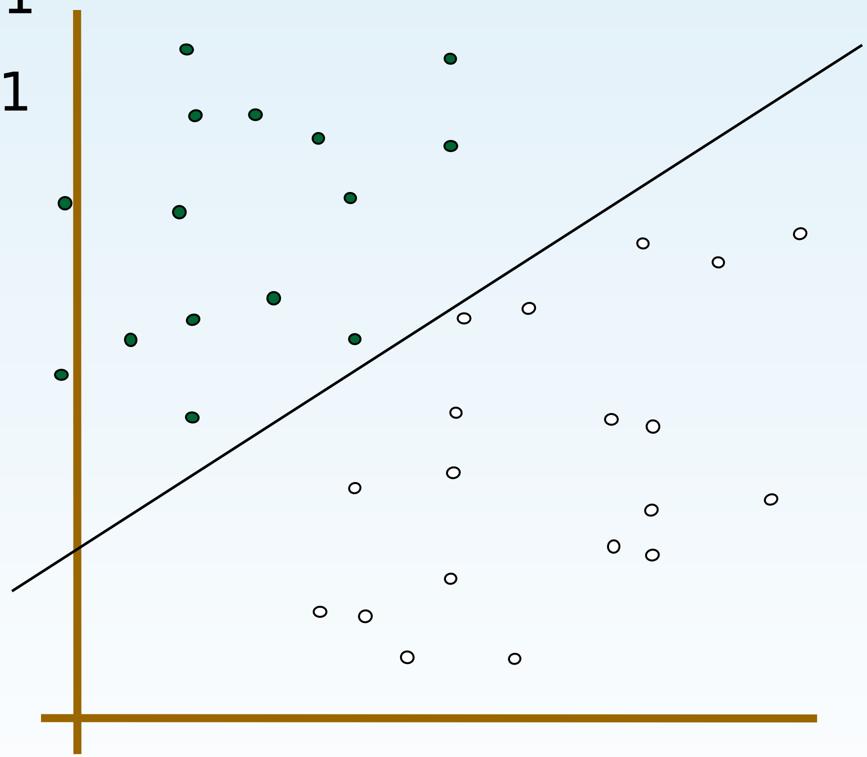
Como classificar esses dados?

De classificadores lineares por funções discriminantes



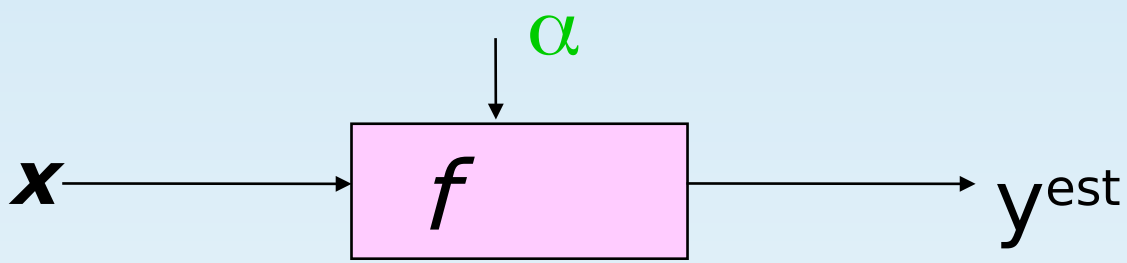
$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w}, b) = sign(\boldsymbol{w}\,\boldsymbol{x} + b)$$

- • indica +1
- ○ indica -1



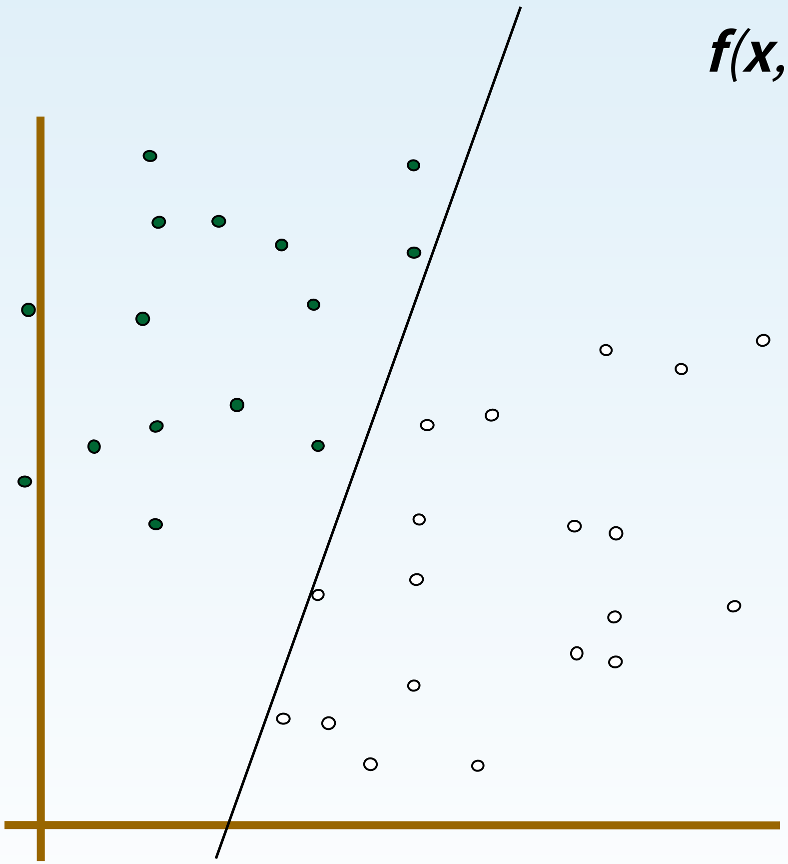
Como classificar esses dados?

De classificadores lineares por funções discriminantes



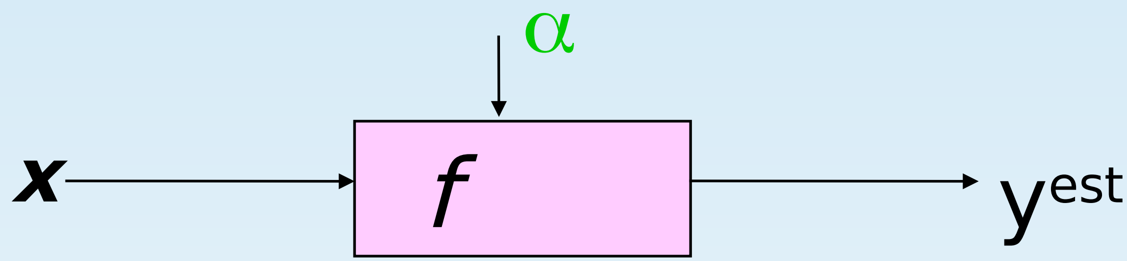
$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}\,\mathbf{x} + b)$$

- ● indica +1
- ○ indica -1



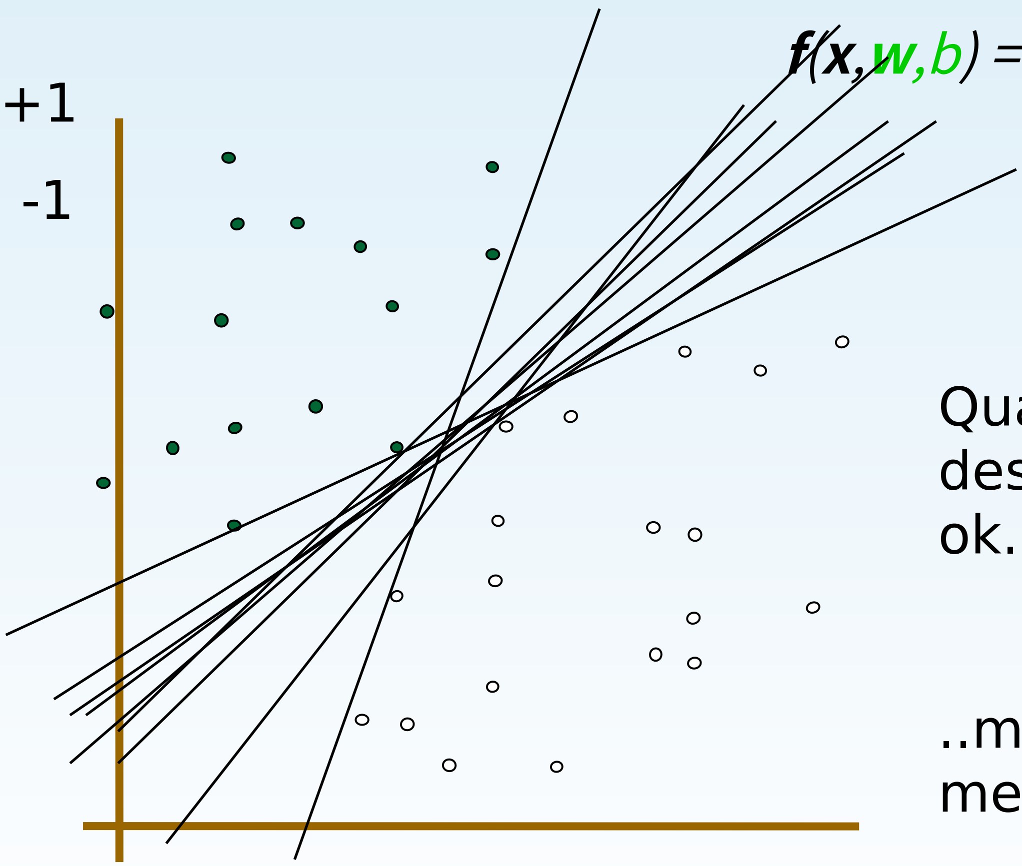
Como classificar esses dados?

De classificadores lineares por funções discriminantes



$f(\mathbf{x}, \mathbf{w}, b) = sign(\mathbf{w}\,\mathbf{x} + b)$

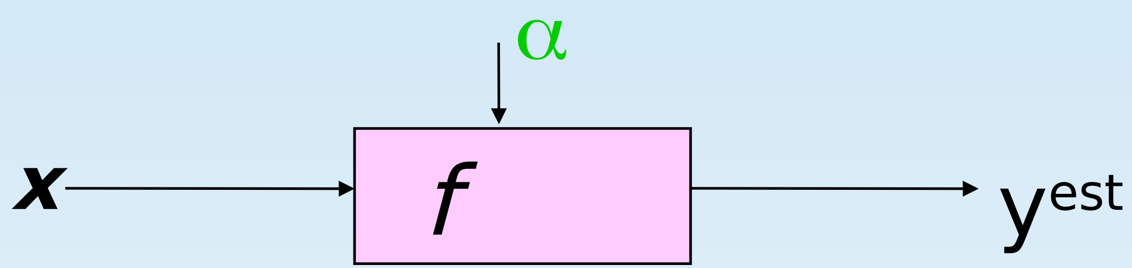
- • indica +1
- ◦ indica -1



Qualquer destas seria ok...

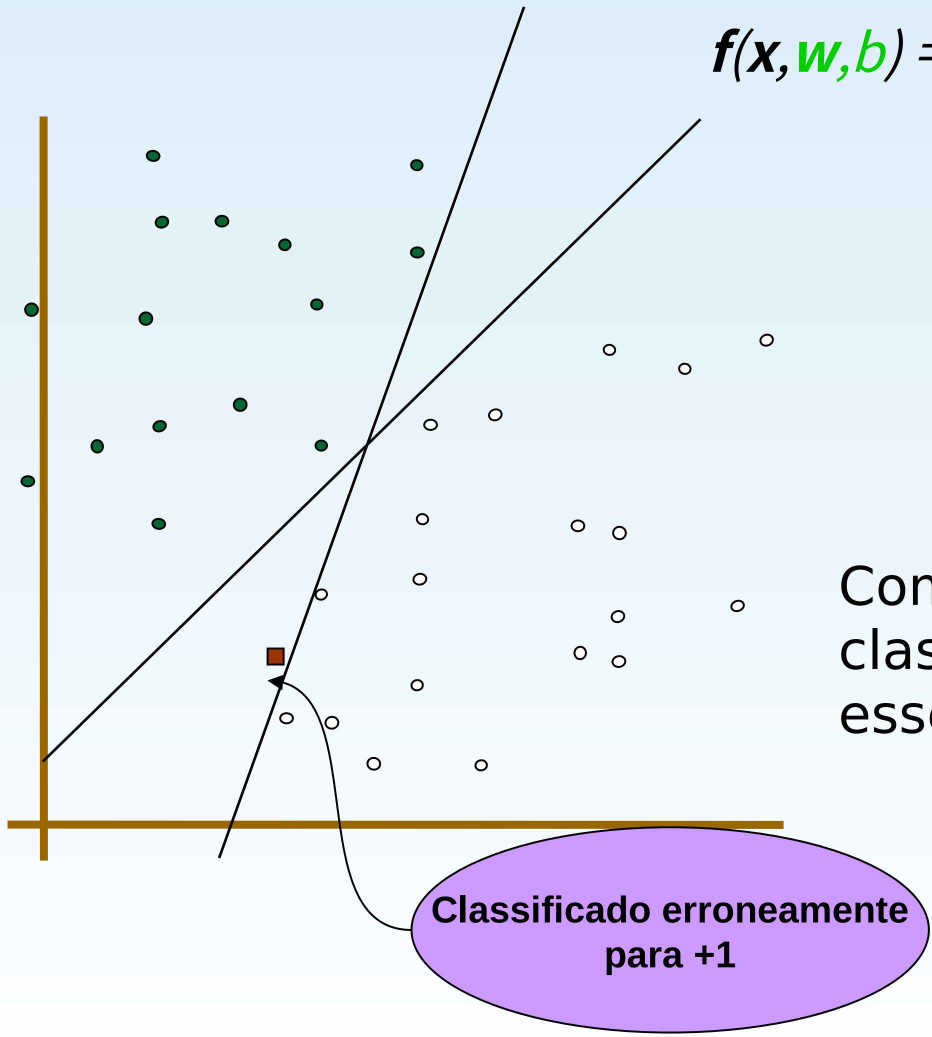
..mas qual é melhor?

De classificadores lineares por funções discriminantes



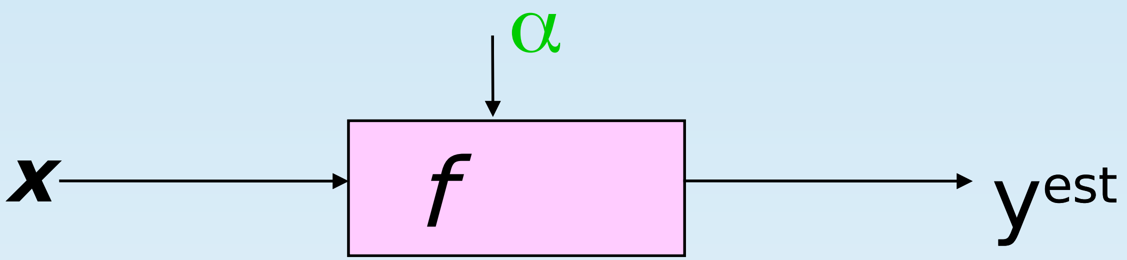
$$f(x, w, b) = sign(w\,x + b)$$

- indica +1
- indica -1

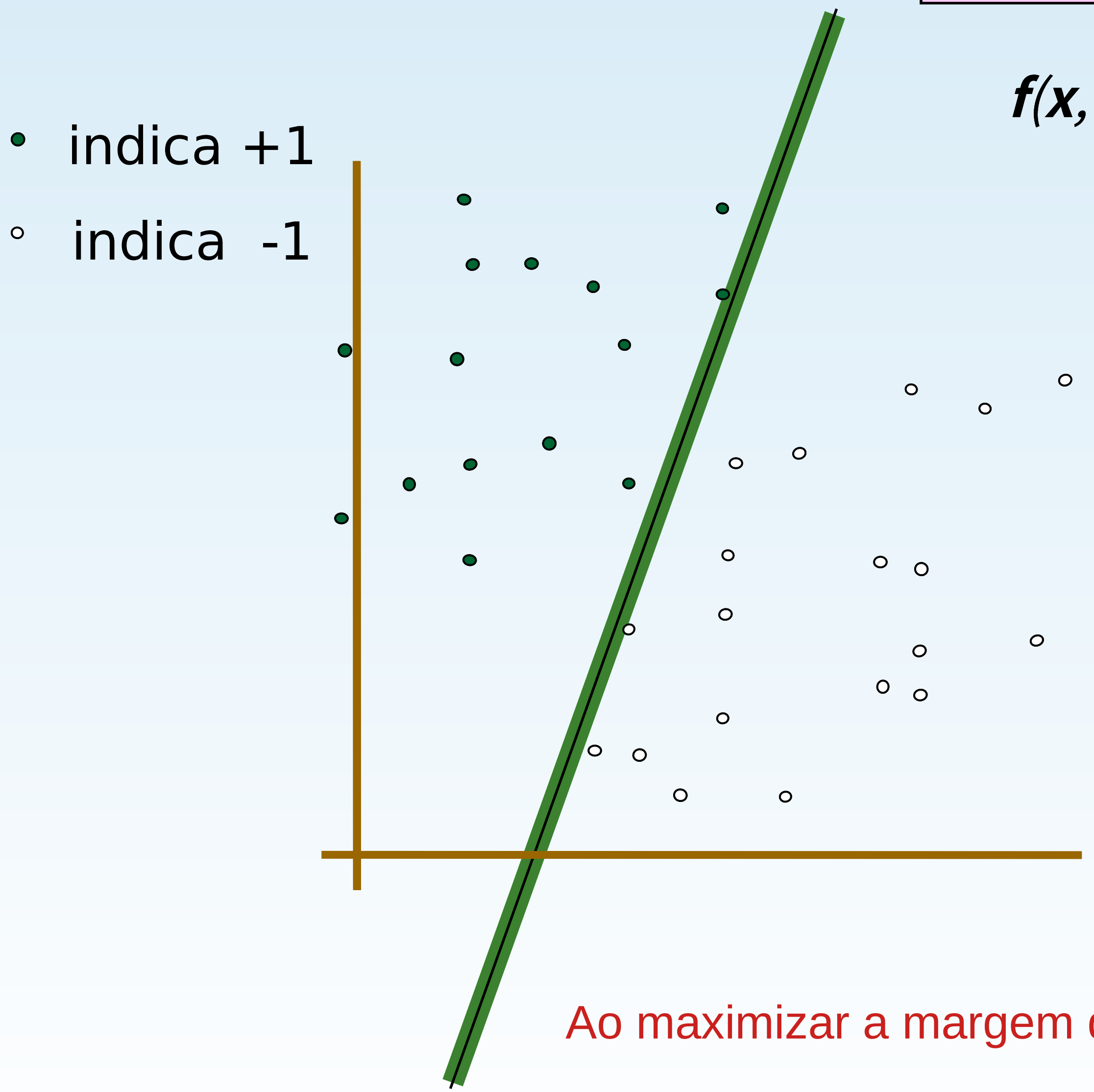


Como classificar esses dados?

Margem



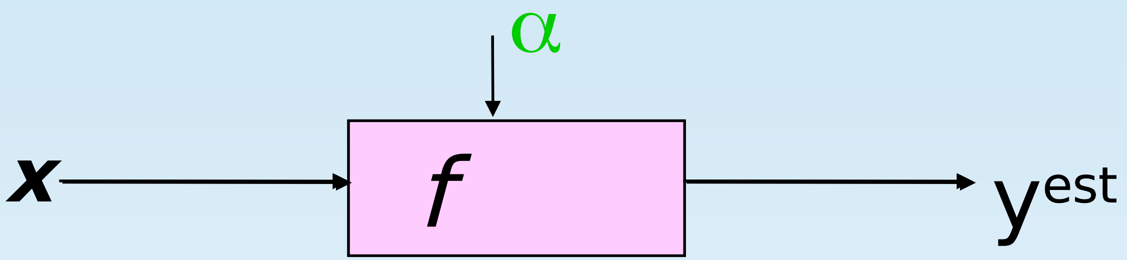
$f(\mathbf{x}, \mathbf{w}, b) = sign(\mathbf{w}\,\mathbf{x} + b)$



Definir a margem de um classificador linear como a largura que o contorno poderia ser aumentado até encontrar um dado(ponto).

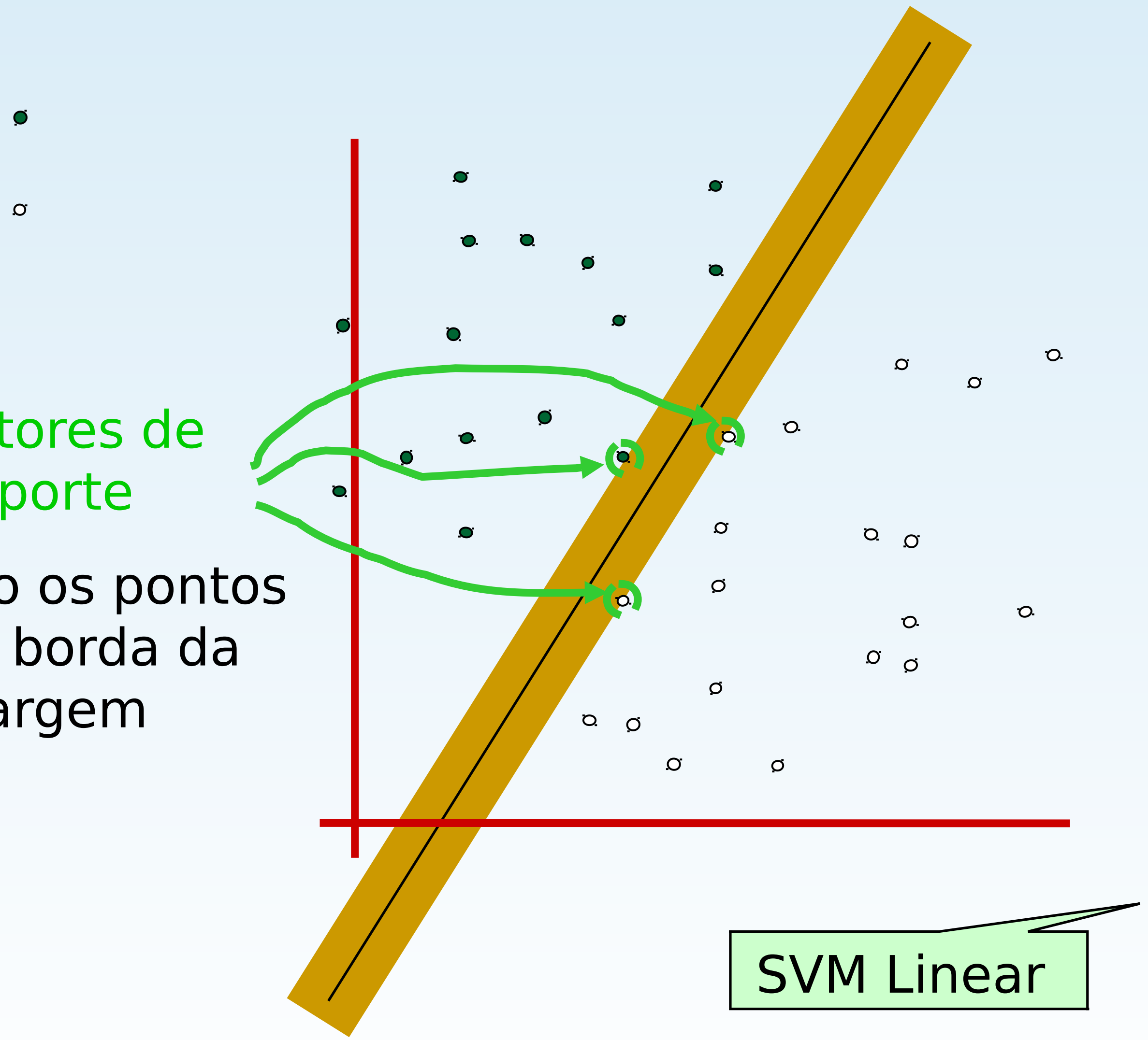
Ao maximizar a margem diminuem-se os modelos possíveis

Margem Máxima

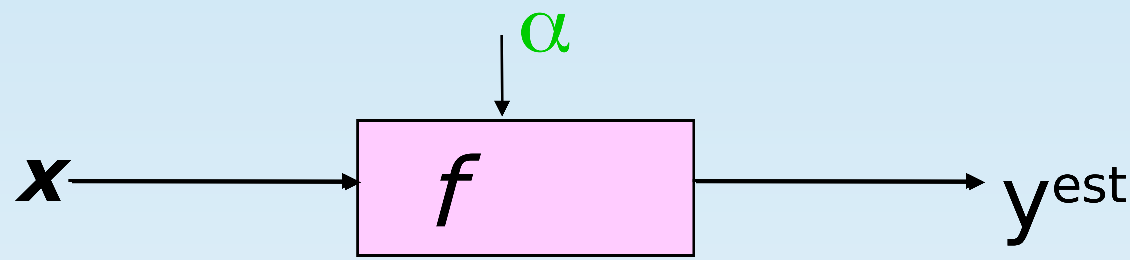


Vetores de Suporte

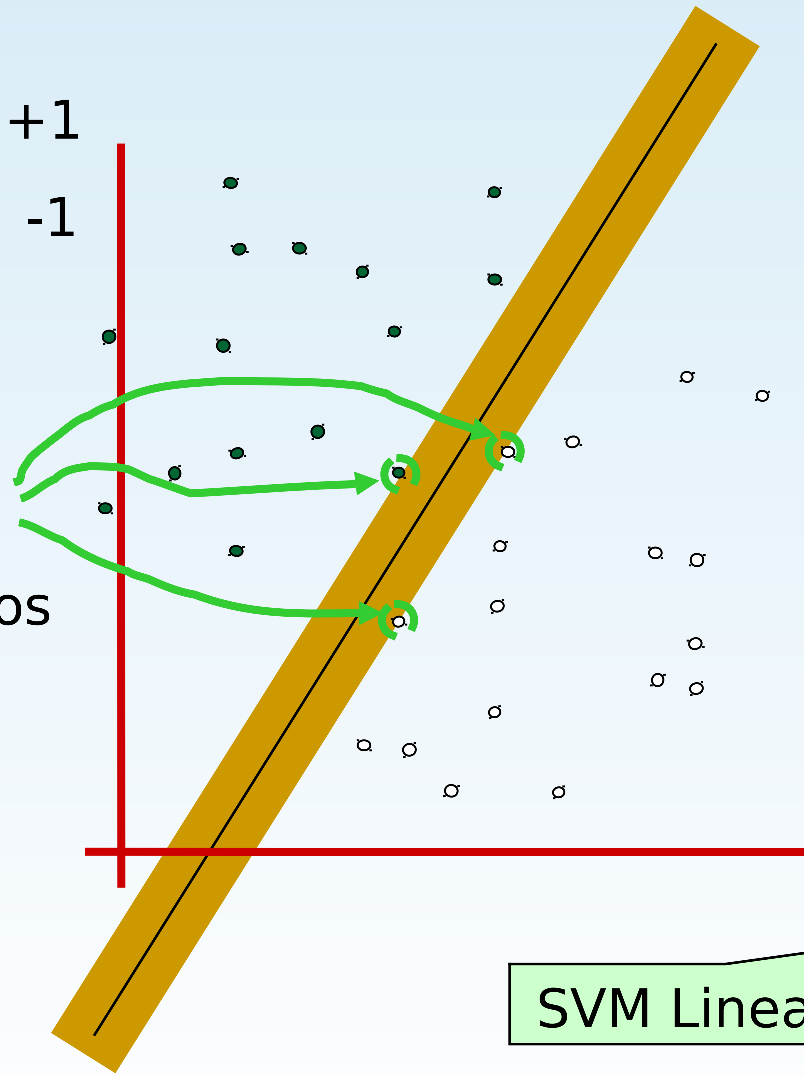
são os pontos na borda da margem



Margem Máxima



- indica +1
- indica -1



Vetores de Suporte

são os pontos na borda da margem

1. Maximizar a margem é uma boa ideia.
2. Implica que somente vetores de suporte são importantes; outros elementos no treinamento são ignoráveis.
3. Empiricamente funciona muito bem.

SVM Linear algoritmo

- Meta: **1) Corretamente classificar todos os dados de treinamento**

$$wx_i + b \geq 1 \qquad \textit{if } y_i = +1$$
$$wx_i + b \leq 1 \qquad \textit{if } y_i = -1$$
$$y_i(wx_i + b) \geq 1 \qquad \text{for all i}$$

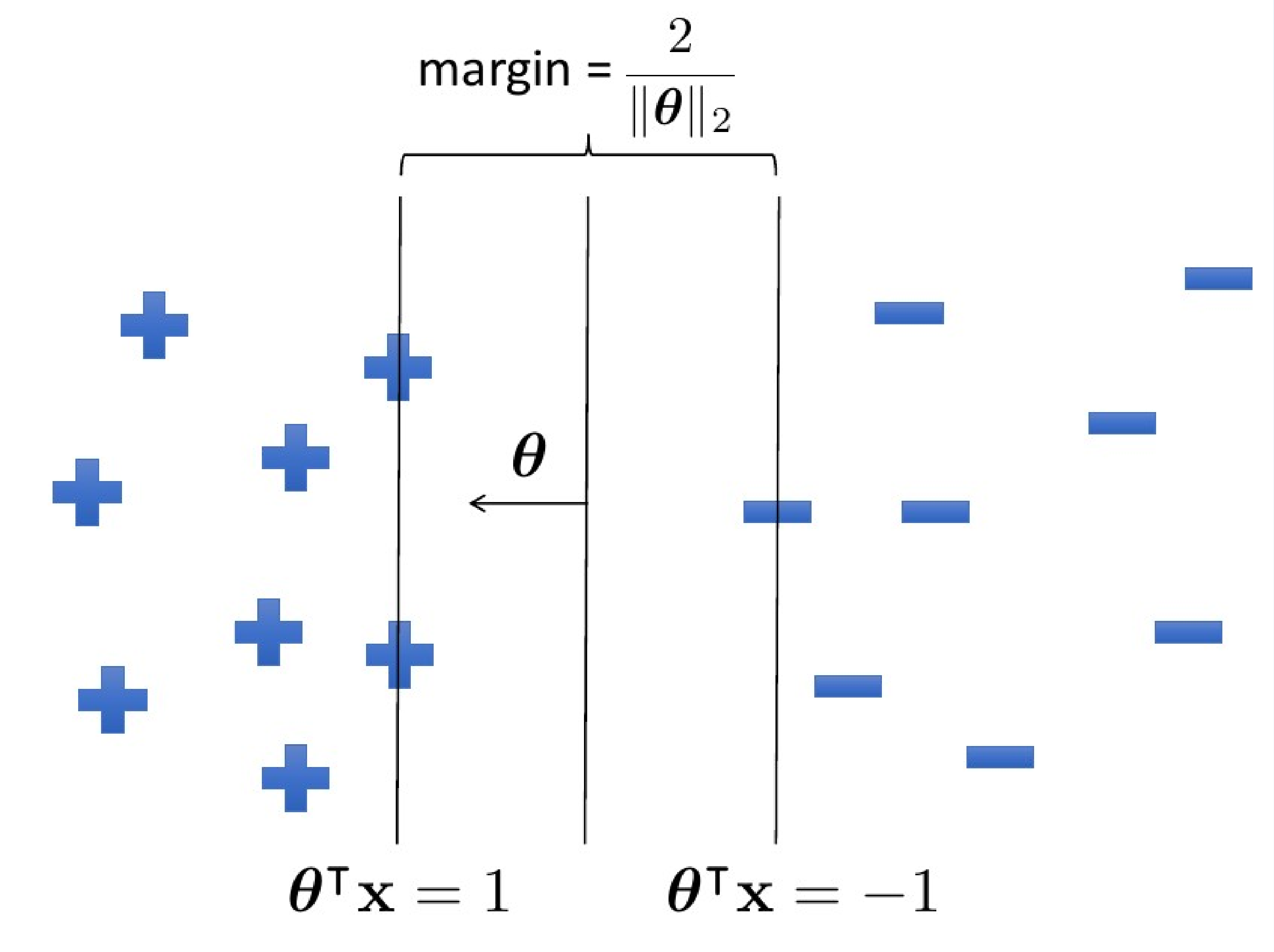
2) Maximizar a Margem $M = \frac{2}{|w|}$

mesmo que minimizar $\frac{1}{2} w^t w$

- **Podemos formular um Problema de Otimização Quadrática e resolver para w e b**
- Minimizar $\Phi(w) = \frac{1}{2} w^t w$

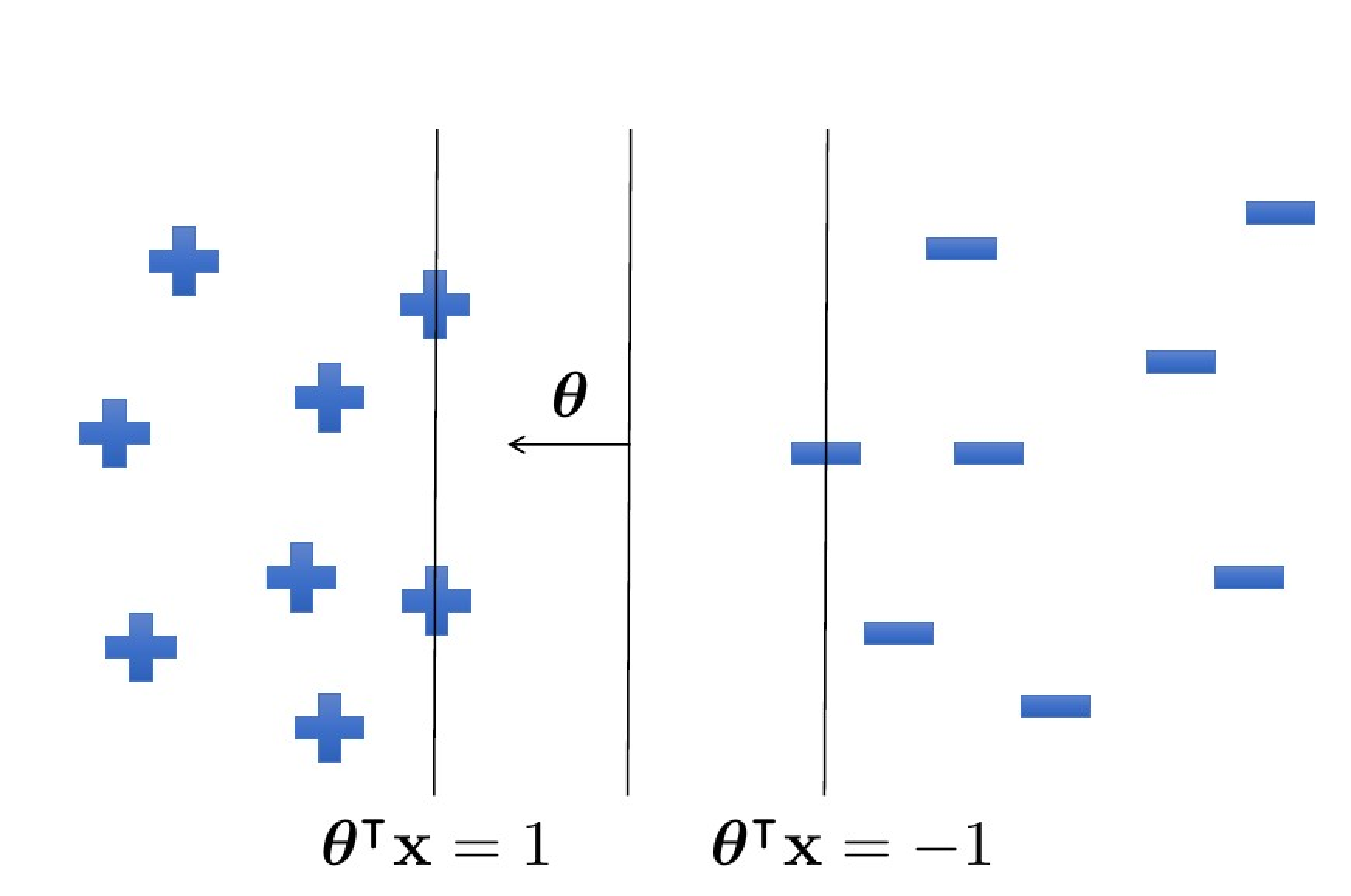
sujeito a $y_i(wx_i + b) \geq 1 \quad \forall i$

Hiperplano Margem Máxima



imagem/slide: E.Eaton

Vetores de Suporte



imagem/slide: E.Eaton

Hiperplano Separador Ótimo

$$\mathcal{X} = \{\mathbf{x}^t, r^t\}_t \text{ where } r^t = \begin{cases} +1 & \text{if } \mathbf{x}^t \in C_1 \\ -1 & \text{if } \mathbf{x}^t \in C_2 \end{cases}$$

find $\mathbf{w}$ and w_0 such that

$$\mathbf{w}^T\mathbf{x}^t + w_0 \geq +1 \text{ for } r^t = +1$$

$$\mathbf{w}^T\mathbf{x}^t + w_0 \leq +1 \text{ for } r^t = -1$$

which can be rewritten as

$$r^t\left(\mathbf{w}^T\mathbf{x}^t + w_0\right) \geq +1$$

(Cortes and Vapnik, 1995; Vapnik, 1995)

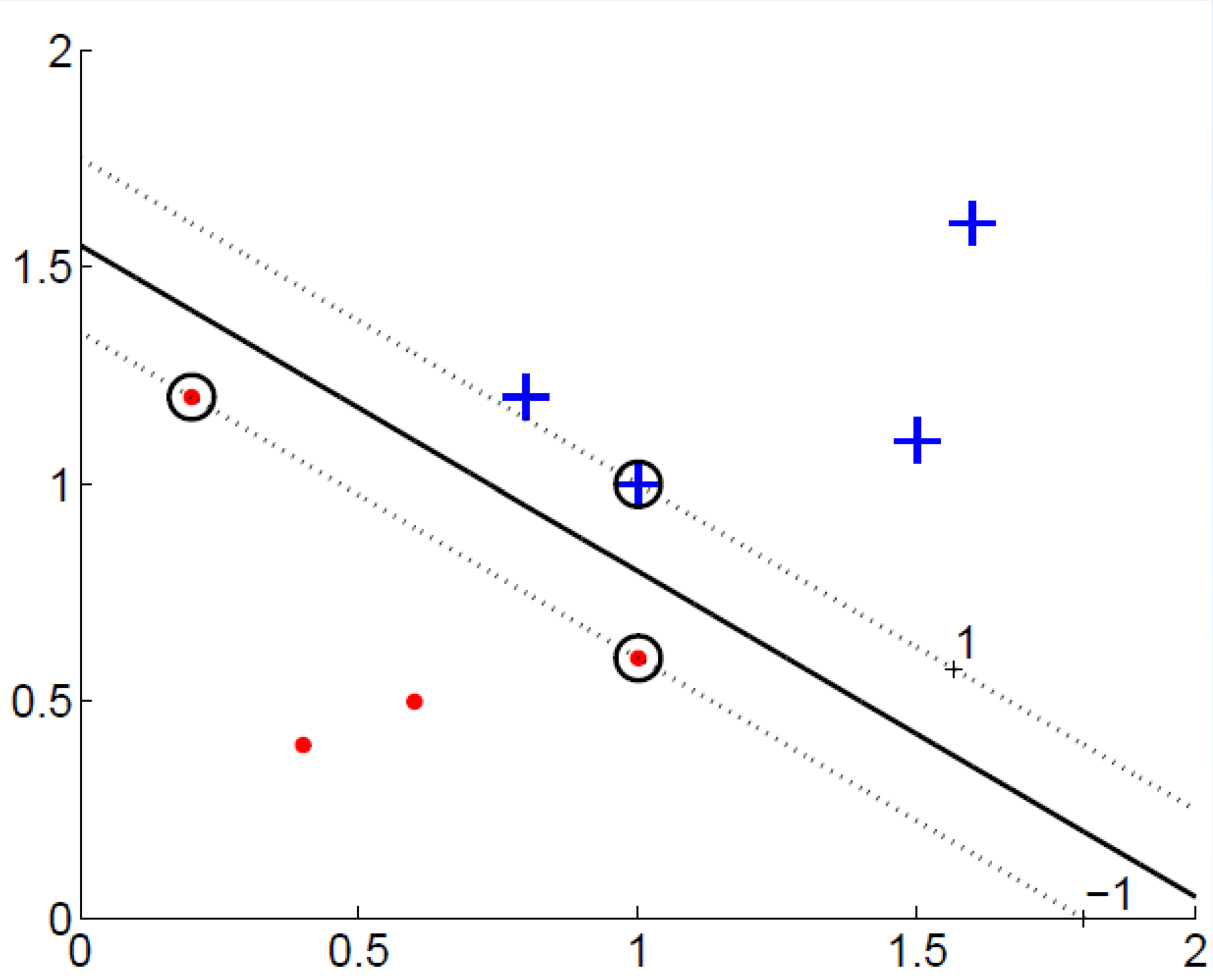
Margem

- Distância do discriminante às instâncias mais próximas em ambos os lados
- Distância de x para o hiperplano é $\dfrac{\left|\mathbf{w}^T\mathbf{x}^t + w_0\right|}{\|\mathbf{w}\|}$
- Requer-se $\dfrac{r^t\left(\mathbf{w}^T\mathbf{x}^t + w_0\right)}{\|\mathbf{w}\|} \geq \rho, \forall t$
- Para uma solução única, fixar $\rho\|\mathbf{w}\|=1$, e para margem máxima

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \text{ subject to } r^t\left(\mathbf{w}^T\mathbf{x}^t + w_0\right) \geq +1, \forall t$$

Classificador Vetor de Suporte (SVM)

Margem



Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)

16

Classificador Vetor de Suporte (SVM)

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \text{ subject to } r^t\left(\mathbf{w}^T\mathbf{x}^t + w_0\right) \geq +1, \forall t$$

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{t=1}^{N} \alpha^t \left[r^t \left(\mathbf{w}^T\mathbf{x}^t + w_0\right) - 1 \right]$$

$$= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{t=1}^{N} \alpha^t r^t \left(\mathbf{w}^T\mathbf{x}^t + w_0\right) + \sum_{t=1}^{N} \alpha^t$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{t=1}^{N} \alpha^t r^t \mathbf{x}^t$$

$$\frac{\partial L_p}{\partial w_0} = 0 \Rightarrow \sum_{t=1}^{N} \alpha^t r^t = 0$$

Classificador Vetor de Suporte (SVM)

$$L_d = \frac{1}{2}\left(\mathbf{w}^T\mathbf{w}\right) - \mathbf{w}^T \sum_t \alpha^t r^t \mathbf{x}^t - w_0 \sum_t \alpha^t r^t + \sum_t \alpha^t$$

$$= -\frac{1}{2}\left(\mathbf{w}^T\mathbf{w}\right) + \sum_t \alpha^t$$

$$= -\frac{1}{2}\sum_t \sum_s \alpha^t \alpha^s r^t r^s \left(\mathbf{x}^t\right)^T \mathbf{x}^s + \sum_t \alpha^t$$

subject to $\sum_{\mathrm{t}} \alpha^t r^t = 0$ and $\alpha^t \geq 0, \forall t$

Muitos α^t são 0 e somente um pequeno número tem $\alpha^t > 0$; eles são os vetores de suporte

Hiperplanos Separáveis

- Imagine uma situação onde tem-se um problema de classificação com duas classes e dois preditores X1 e X2.

Hiperplanos Separáveis

- Imagine uma situação onde tem-se um problema de classificação com duas classes e dois preditores X1 e X2.
- Suponha que as duas classes são "separáveis linearmente" i.e. pode-se desenhar uma reta onde todos os pontos de um lado pertencem à primeira classe, e pontos do outro lado à segunda classe.

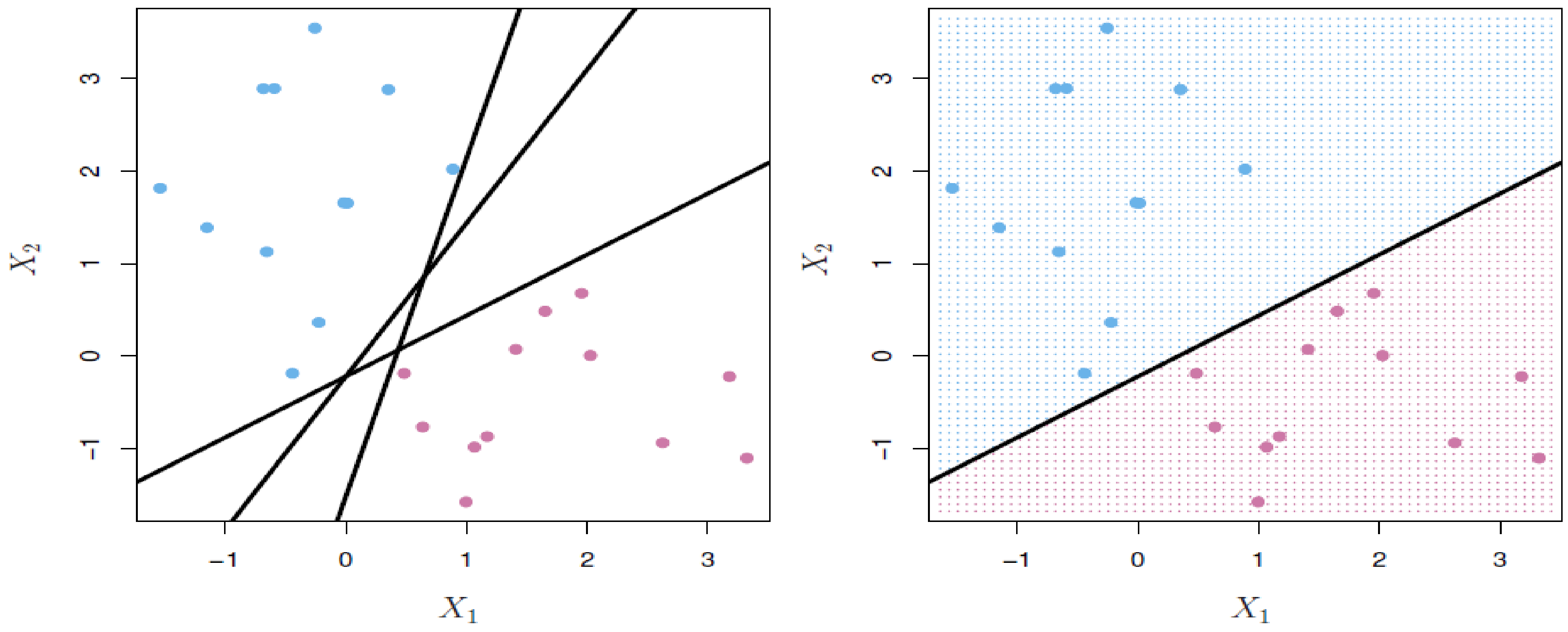
Hiperplanos Separáveis

- Imagine uma situação onde tem-se um problema de classificação com duas classes e dois preditores X1 e X2.
- Suponha que as duas classes são "separáveis linearmente" i.e. pode-se desenhar uma reta onde todos os pontos de um lado pertencem à primeira classe, e pontos do outro lado à segunda classe.
- Então uma abordagem direta é encontrar uma linha reta que forneça a maior separação entre as classes i.e. os pontos estão o mais distante da linha quanto possível.

Hiperplanos Separáveis

- Imagine uma situação onde tem-se um problema de classificação com duas classes e dois preditores X1 e X2.
- Suponha que as duas classes são “separáveis linearmente” i.e. pode-se desenhar uma reta onde todos os pontos de um lado pertencem à primeira classe, e pontos do outro lado à segunda classe.
- Então uma abordagem direta é encontrar uma linha reta que forneça a maior separação entre as classes i.e. os pontos estão o mais distante da linha quanto possível.
- Esta é a ideia básica de um classificador vetor de suporte.

Hiperplanos Separáveis



- If $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, then $f(X) > 0$ for points on one side of the hyperplane, and $f(X) < 0$ for points on the other.
- If we code the colored points as $Y_i = +1$ for blue, say, and $Y_i = -1$ for mauve, then if $Y_i \cdot f(X_i) > 0$ for all i, $f(X) = 0$ defines a *separating hyperplane*.

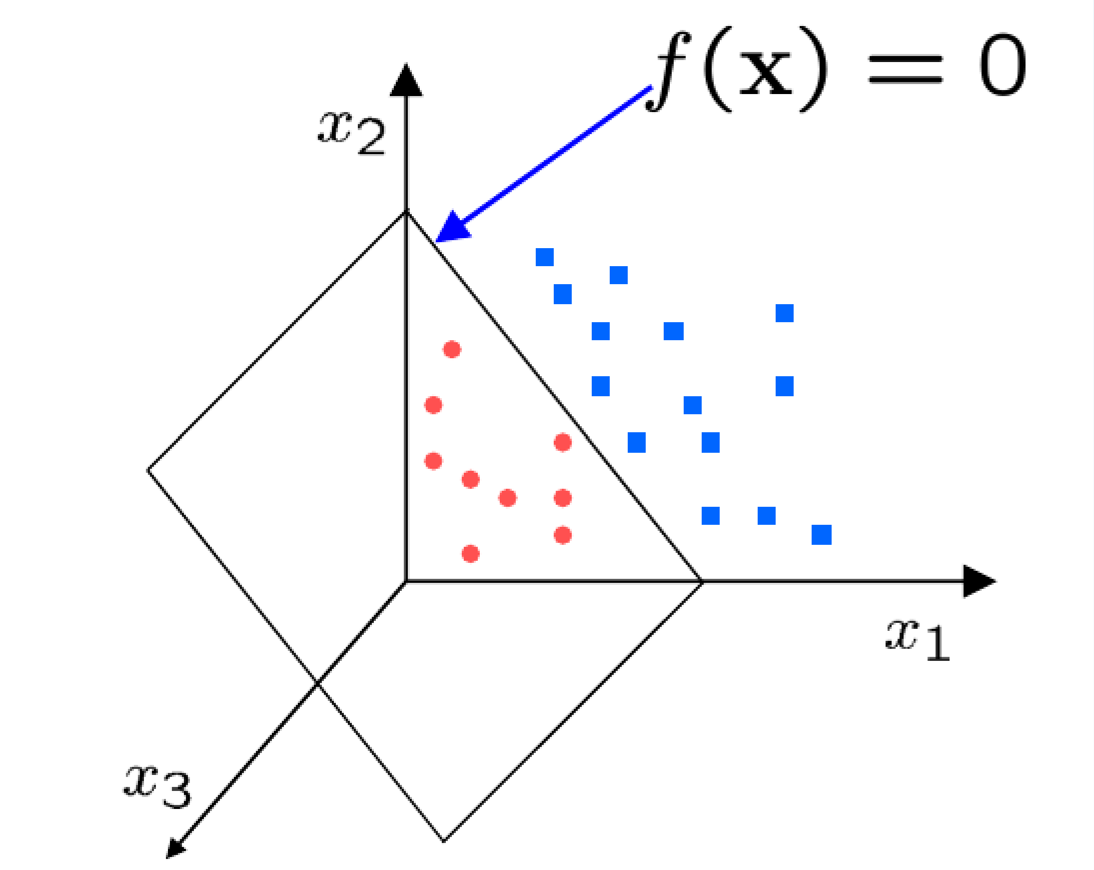
Hiperplanos Separáveis

Equação de um Hiperplano

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = 0$$

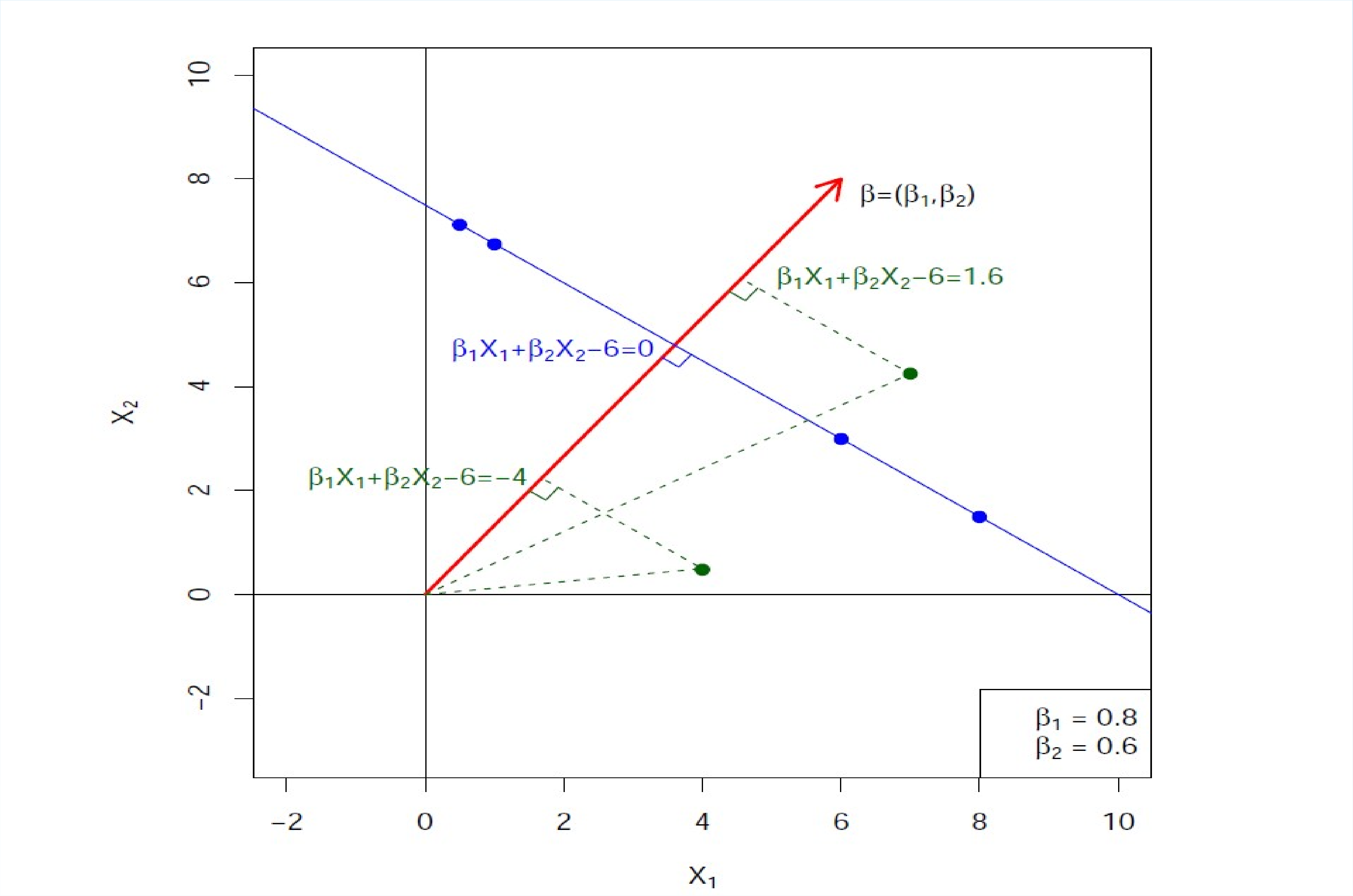
Hiperplanos Separáveis

Exemplo de um Hiperplano



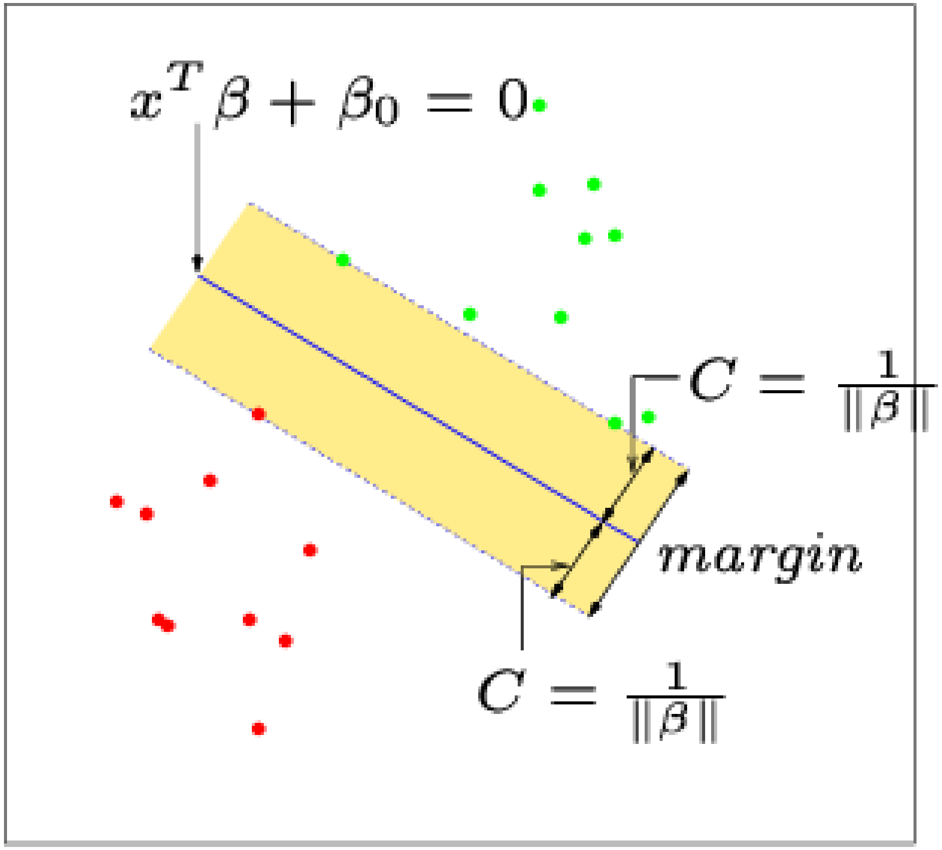
Hiperplanos Separáveis

Hiperplano em 2 dimensões



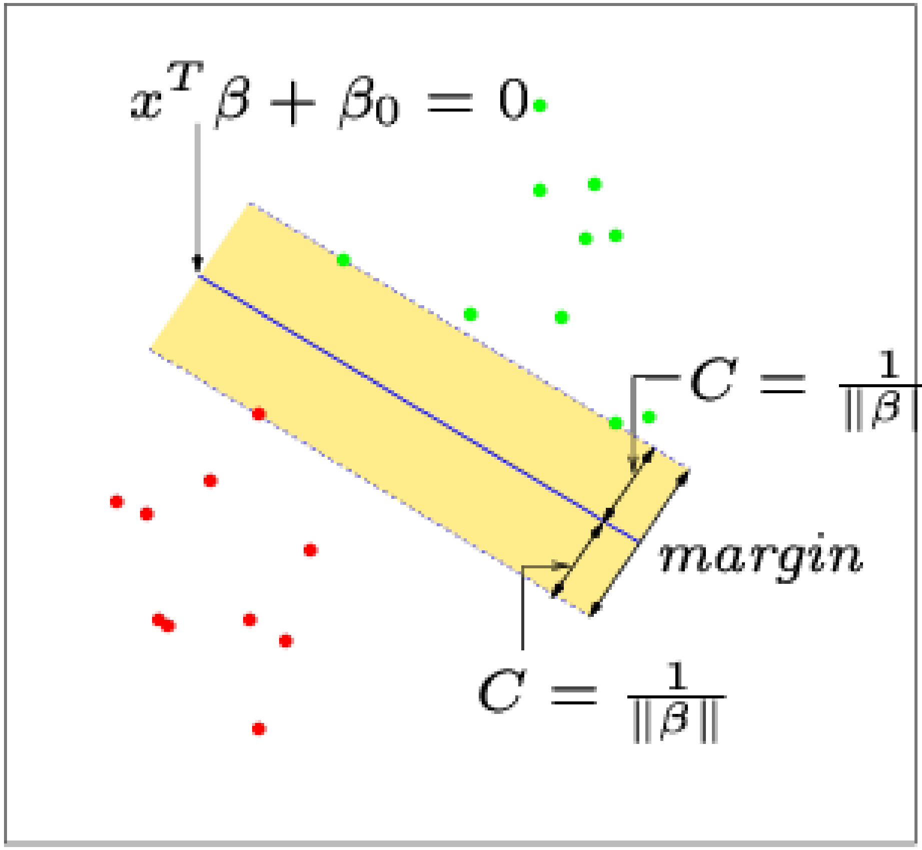
Encontrando Hiperplanos

- C é a distância perpendicular mínima entre cada ponto e a reta separadora.



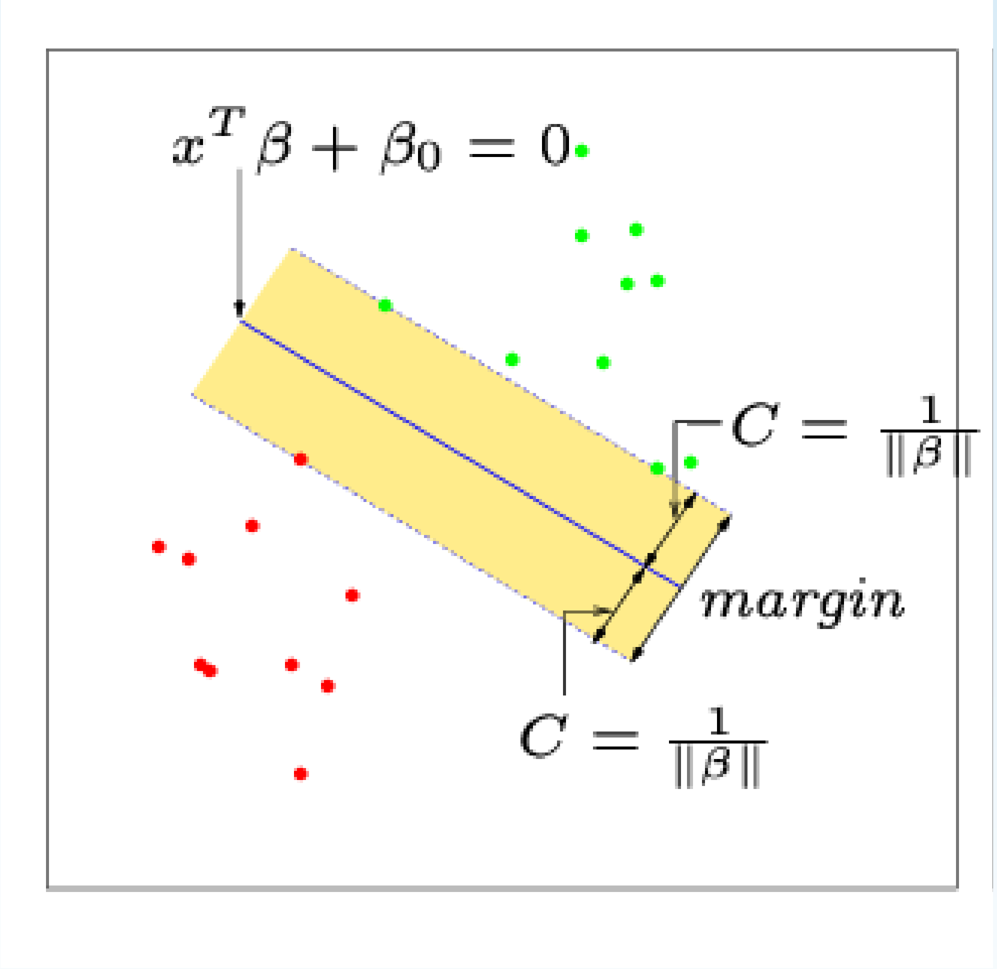
Encontrando Hiperplanos

- C é a distância perpendicular mínima entre cada ponto e a reta separadora.
- Encontramos a reta que maximiza C.



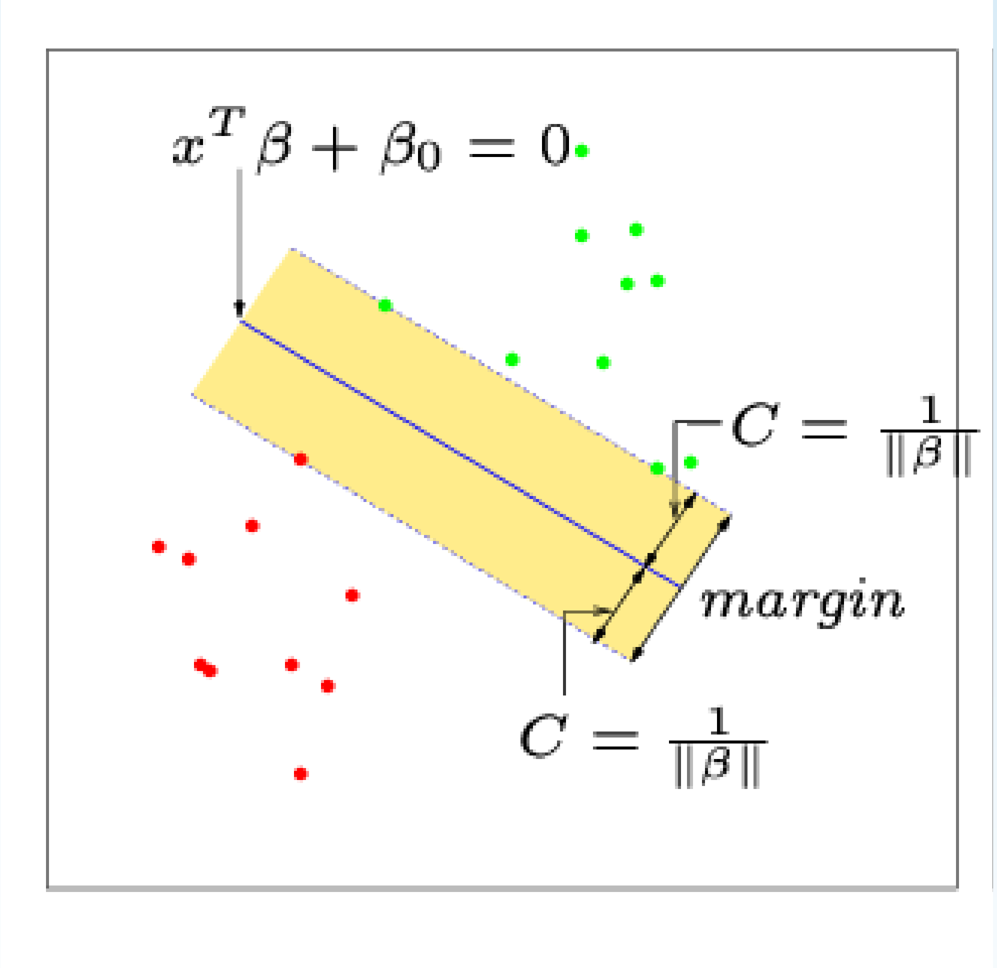
Encontrando Hiperplanos

- C é a distância perpendicular mínima entre cada ponto e a reta separadora.
- Encontramos a reta que maximiza C.
- Essa reta é chamada de “hiperplano separador ótimo”.



Encontrando Hiperplanos

- C é a distância perpendicular mínima entre cada ponto e a reta separadora.
- Encontramos a reta que maximiza C.
- Essa reta é chamada de "hiperplano separador ótimo"
- A classificação de um ponto depende de qual lado da reta ele fica.



Encontrando Hiperplanos

$$\underset{\beta_0,\beta_1,\ldots,\beta_p}{\text{maximize}}\, M$$

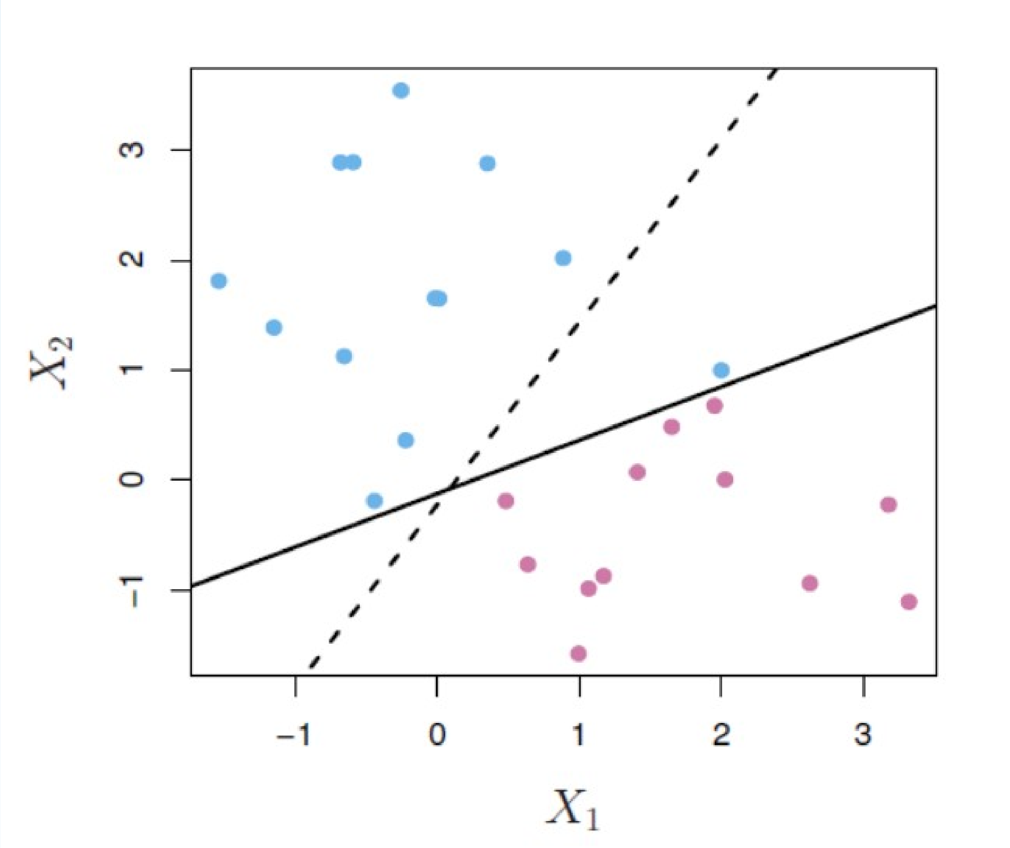
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

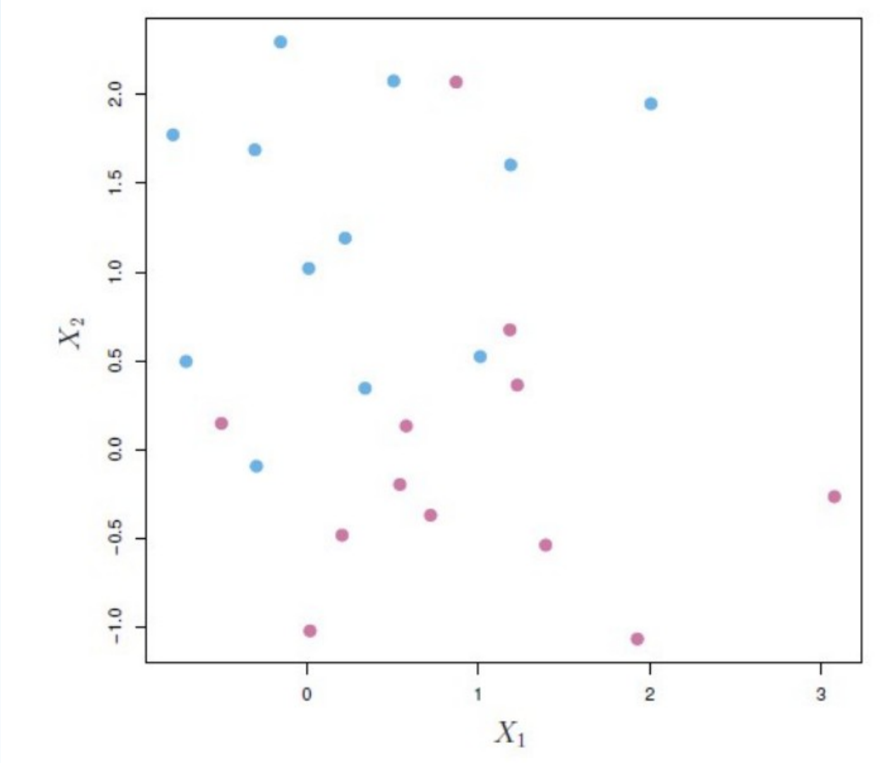
$$y_i(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \geq M$$

$$\text{for all } \; i = 1, \ldots, N.$$

Encontrando Hiperplanos

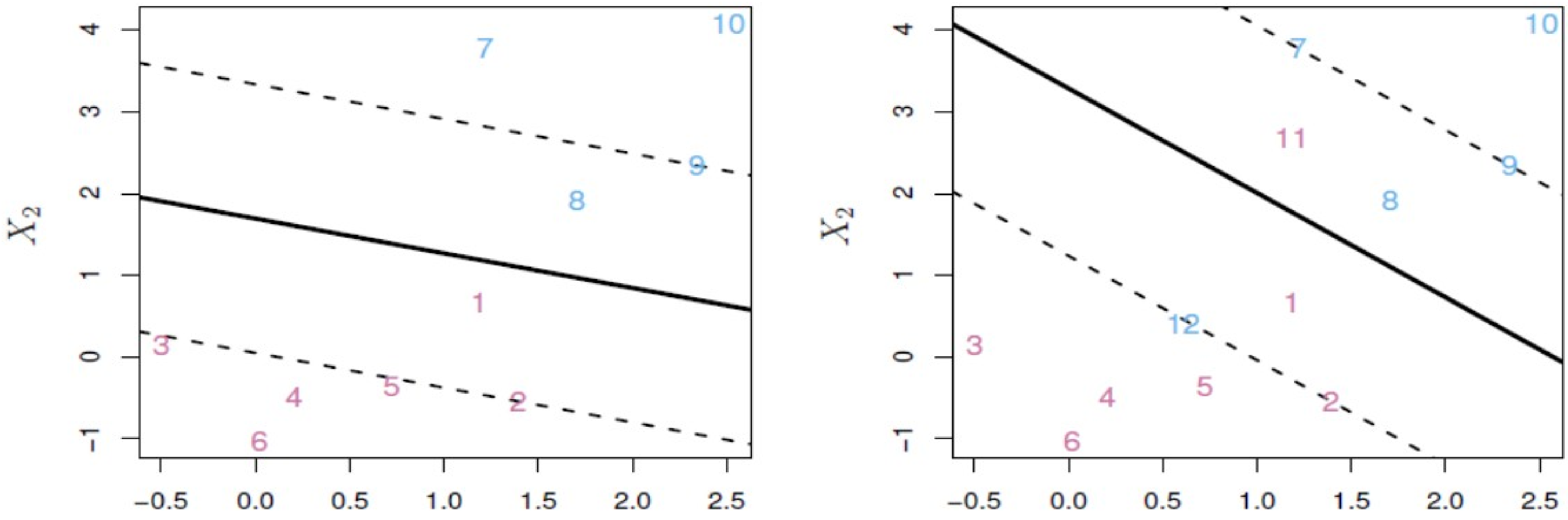
Mas, os dados podem ser ruidosos, ou não haver separabilidade entre classes





SVM

Usar C como regularizador, parâmetro de equilíbrio encontrando vetores de suporte ao hiperplano



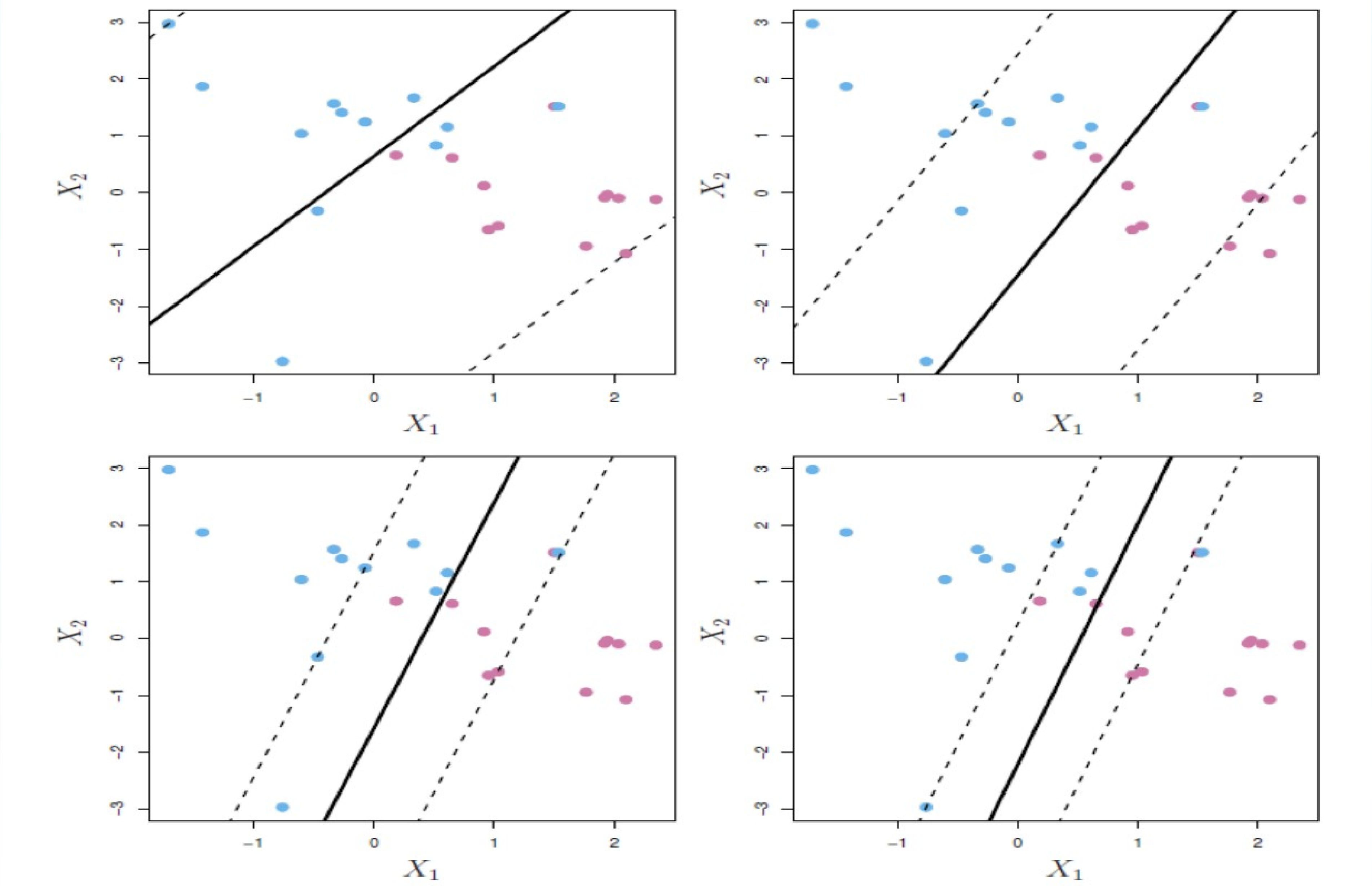
$$\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n}{\text{maximize}} \quad M \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

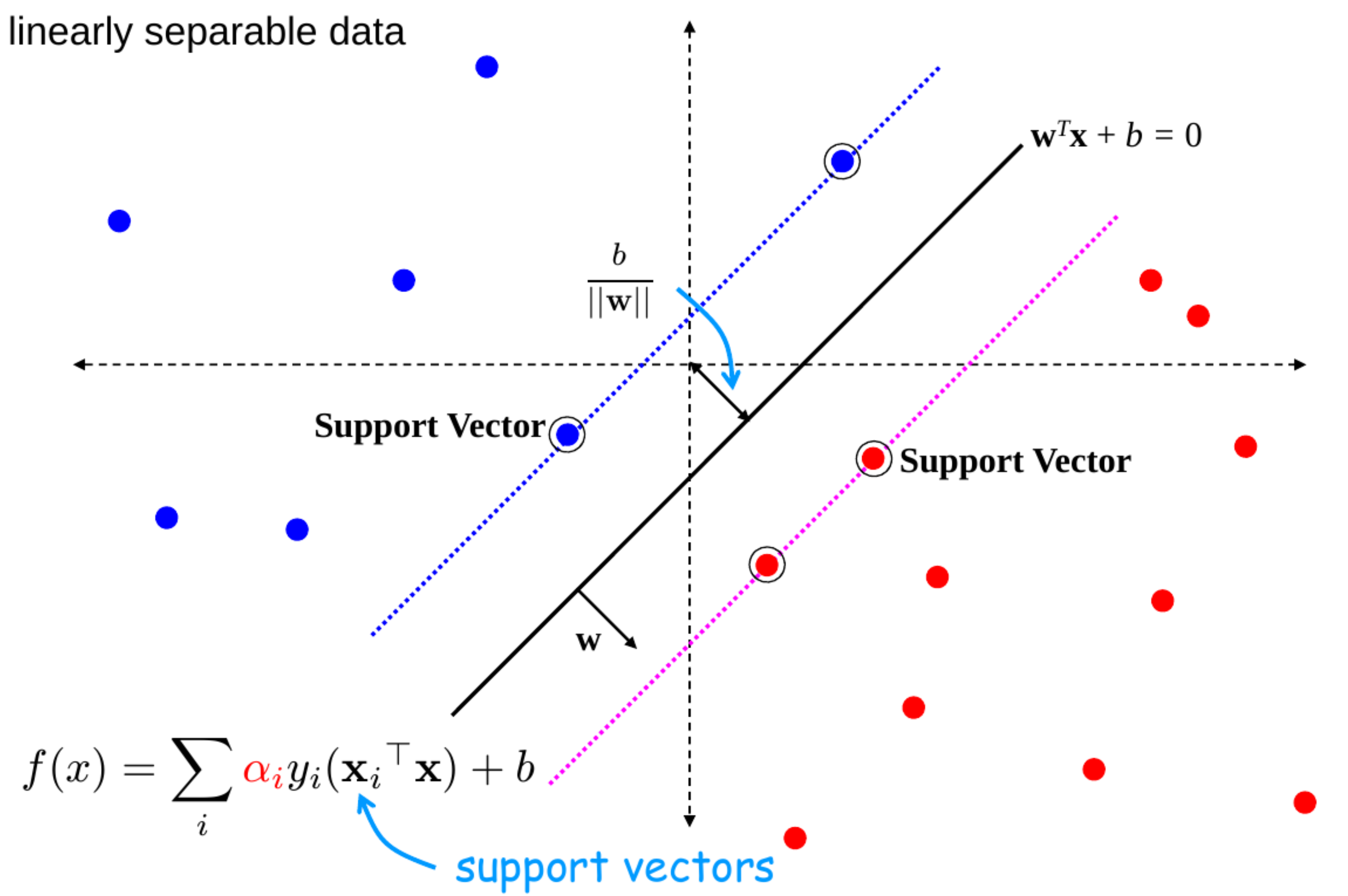
$$\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C,$$

SVM

Usar C como regularizador, parâmetro de equilíbrio encontrando vetores de suporte ao hiperplano



Support Vector Machine



SVM

Classes sem separação exata

- Na prática não é usualmente possível encontrar um hiperplano que separa perfeitamente duas classes.

SVM

Classes sem separação exata

- Na prática não é usualmente possível encontrar um hiperplano que separa perfeitamente duas classes.
- Haverá pelo menos alguns pontos no lado errado da reta.

SVM

Classes sem separação exata

- Na prática não é usualmente possível encontrar um hiperplano que separa perfeitamente duas classes.
- Haverá pelo menos alguns pontos no lado errado da reta.
- Nessa situação tentamos encontrar um plano que dá a melhor separação entre os pontos que são corretamente classificados sujeitos a pontos do lado errado em pequena quantidade.

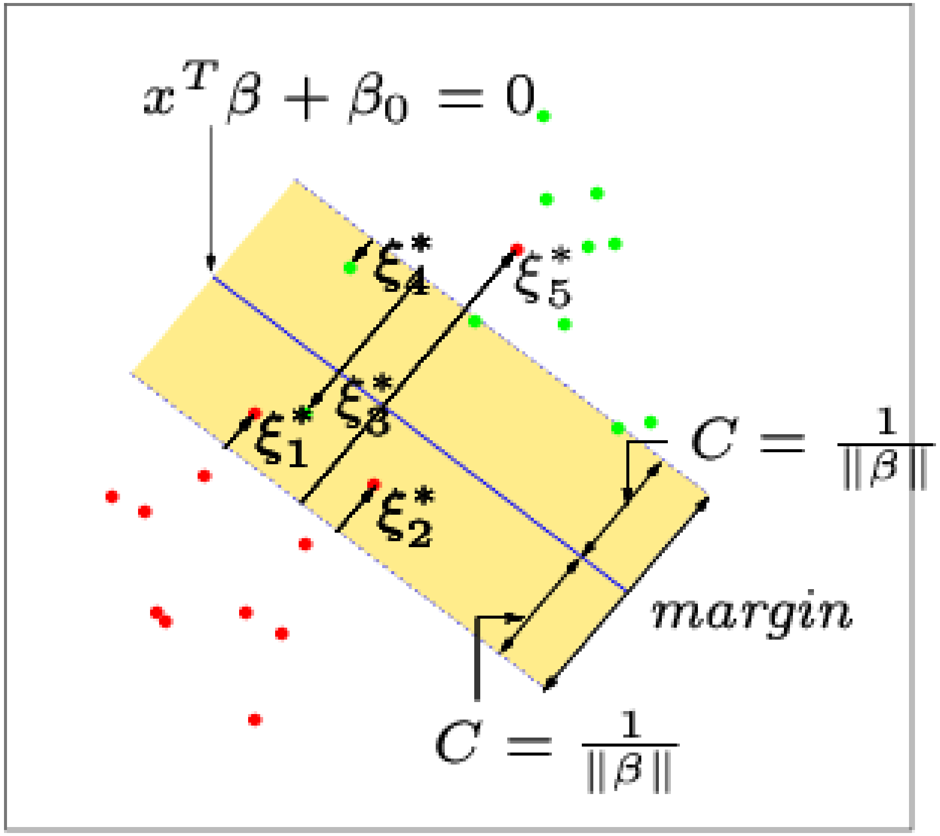
SVM (Flexibilizar Margem)

Variáveis que permitam folga ξi **podem ser adicionados para permitir algumas classificações errôneas, difíceis ou ruídos.**

- Seja ξ*i representando a quantia que o i-ésimo ponto está no lado errado da margem (a reta espaçada).

$$\min_{\boldsymbol{\theta}} \frac{1}{2} \sum_{j=1}^{d} \theta_j^2 + C \sum_i \xi_i$$

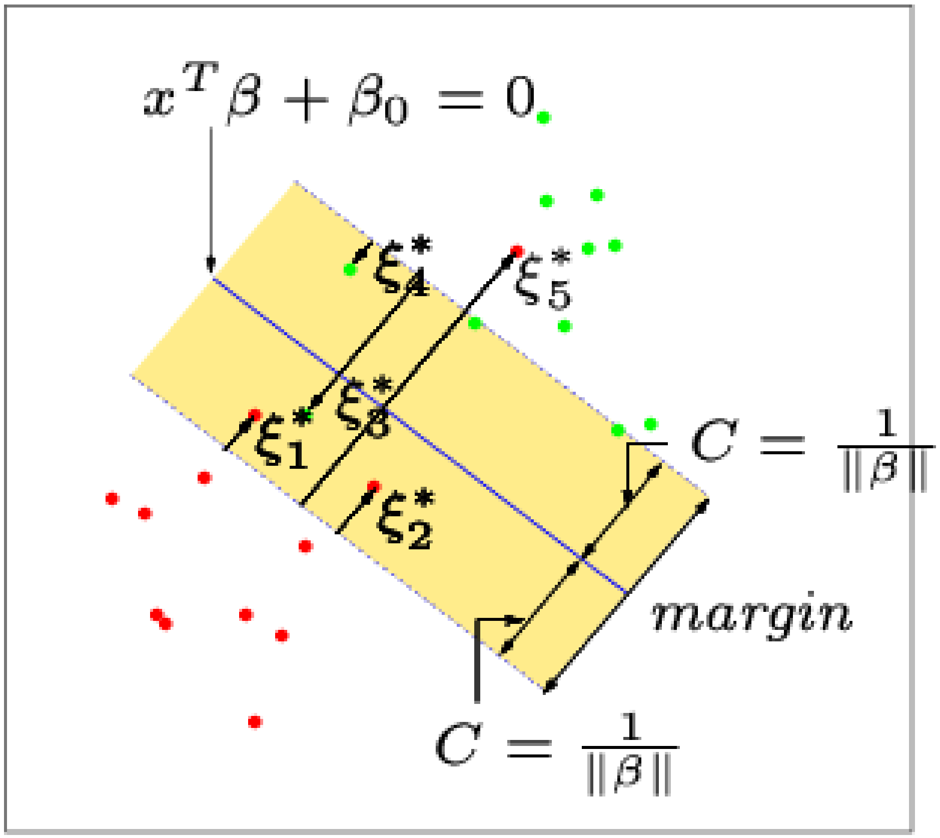
$$\text{s.t.} \quad y_i \left(\boldsymbol{\theta}^\top \boldsymbol{x}_i\right) \geq 1 - \xi_i \quad \forall i$$



SVM (Flexibilizar Margem)

- Seja ξ*i representando a quantia que o i-ésimo ponto está no lado errado da margem (a reta espaçada).
- Então queremos maximizar C sujeito a

$$\frac{1}{C}\sum_{i=1}^{n}\xi_i^* \leq \text{Constant}$$

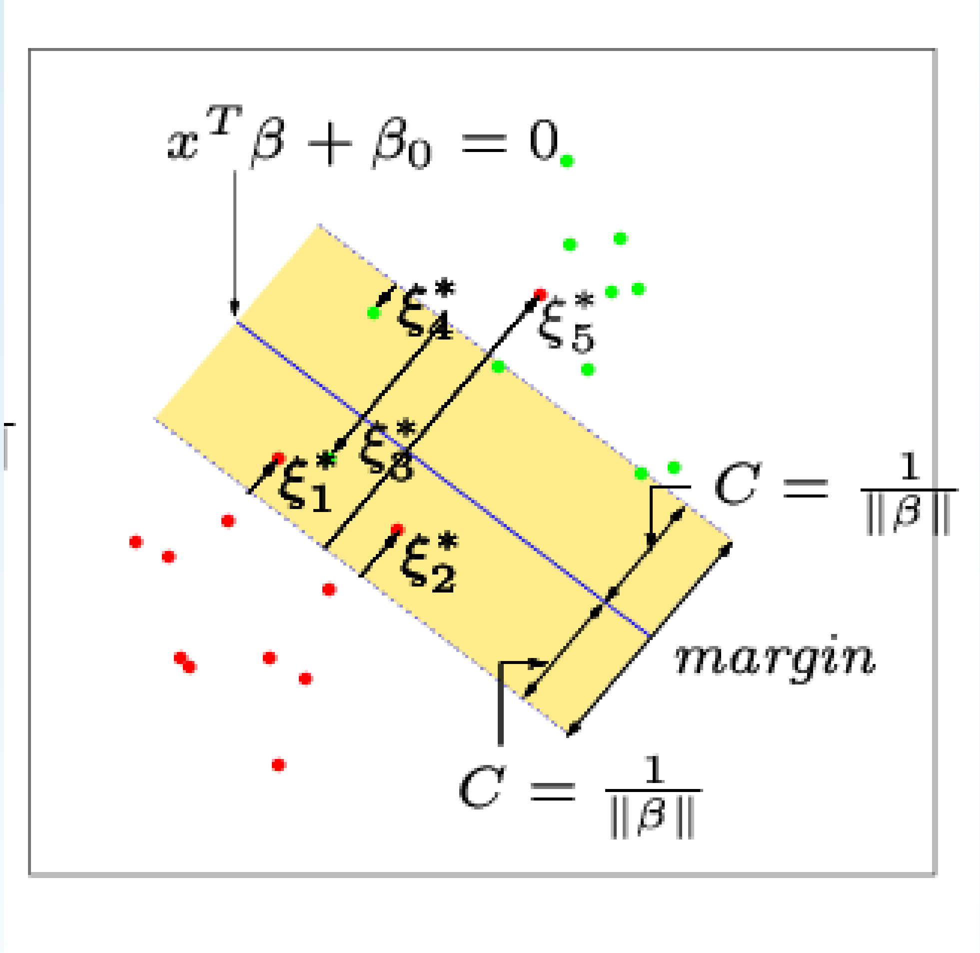


SVM (Flexibilizar Margem)

- Seja ξ*i representando a quantia que o i-ésimo ponto está no lado errado da margem (a reta espaçada).
- Então queremos maximizar C sujeito a

$$\frac{1}{C}\sum_{i=1}^{n}\xi_i^* \leq \text{Constant}$$

- A constante é um parâmetro de ajuste que escolhemos.



Margem Rígida vs. Flexível

- **Rígida:**

Encontrar **w** e ***b*** tais que

$\mathbf{\Phi}(\mathbf{w}) = ½\ \mathbf{w}^T\mathbf{w}$ é minimizada e para todos $\{(\mathbf{x_i}, y_i)\}$

$y_i\ (\mathbf{w^T x_i} + \mathrm{b}) \geq 1$

- **Flexível (incorporando variáveis de folga):**

Encontrar **w** e ***b*** tais que

$\mathbf{\Phi}(\mathbf{w}) = ½\ \mathbf{w}^T\mathbf{w} + C\Sigma\xi_i$ é minimizada e para todos $\{(\mathbf{x_i}, y_i)\}$

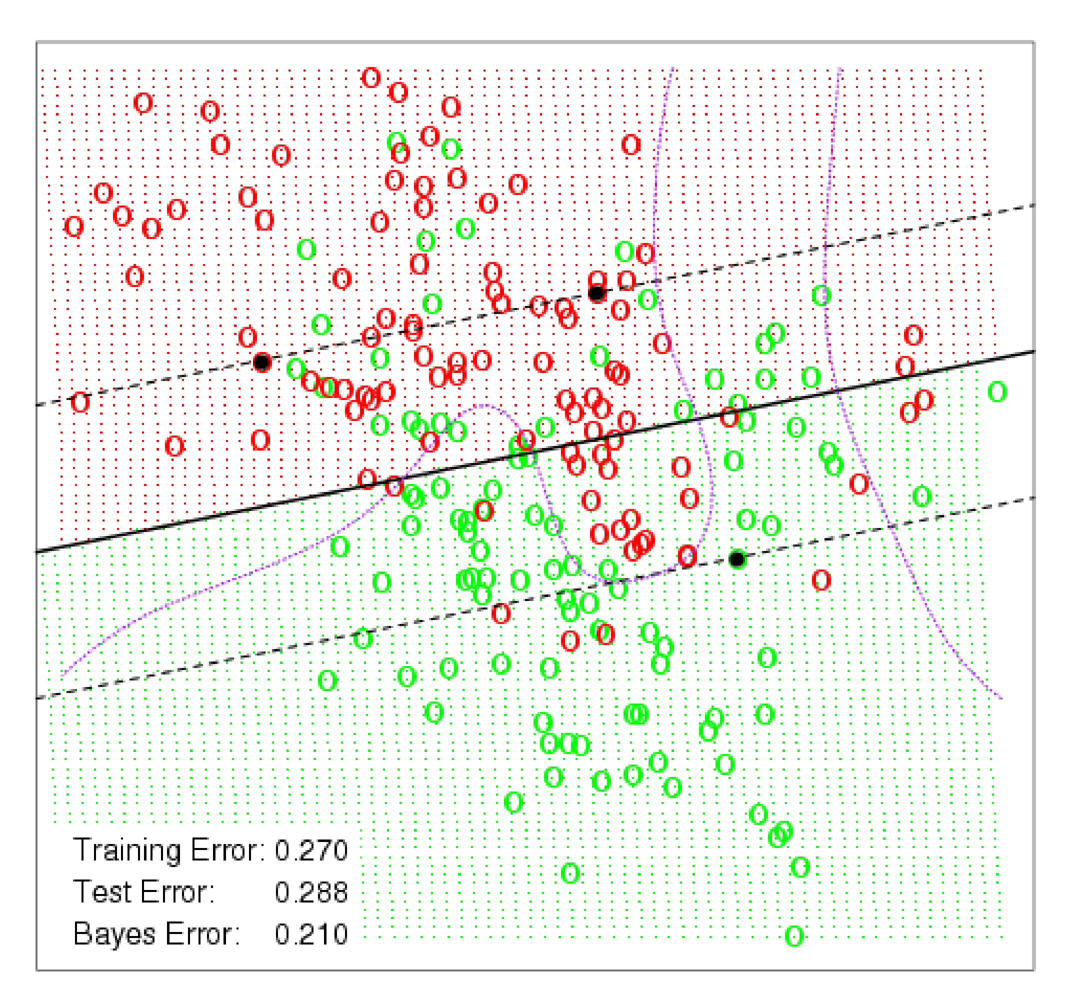
$y_i\ (\mathbf{w^T x_i} + b) \geq 1 - \xi_i$ e $\xi_i \geq 0$ para todos i

- **Parâmetro *C* pode ser visto como uma forma de controlar sobreajuste.**

SVM

Exemplo com constante pequena

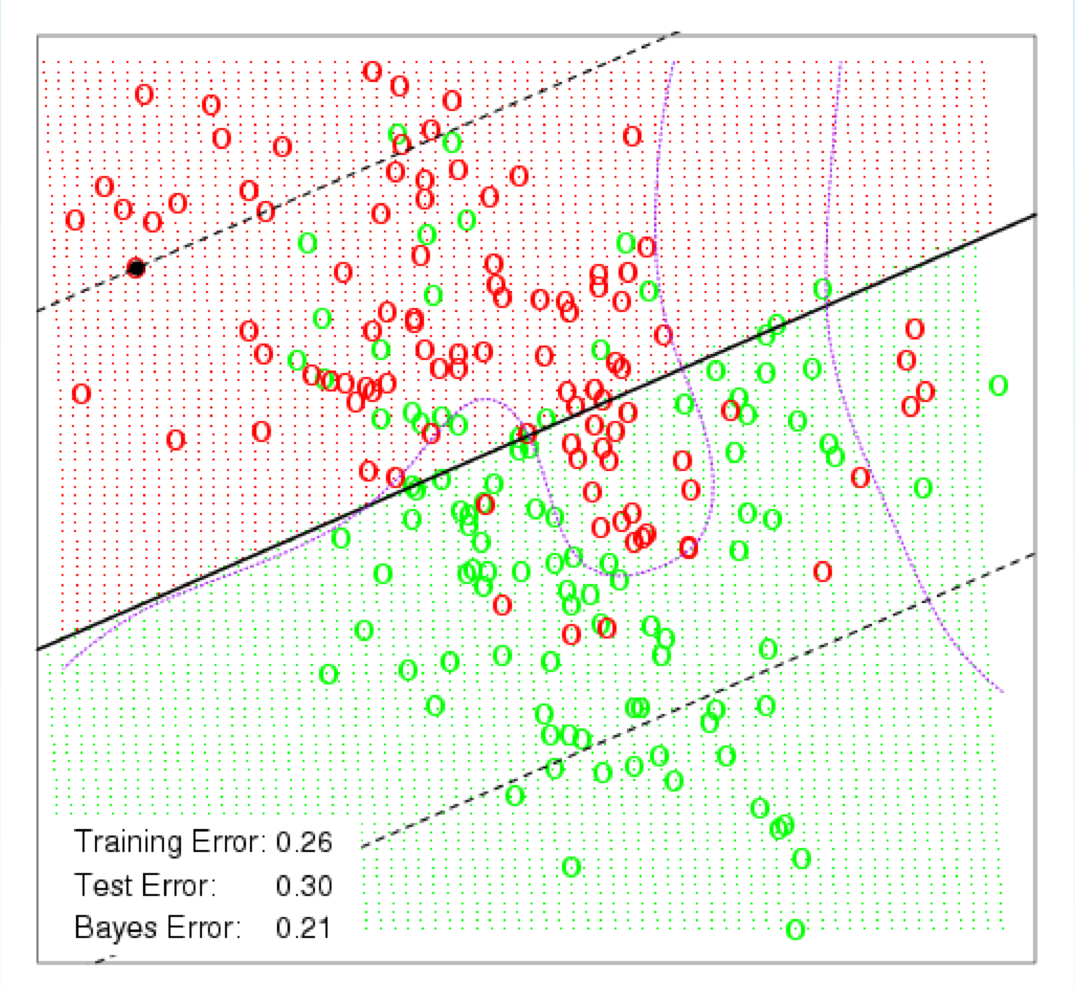
- A distância entre as retas representa a margem ou 2C.
- As curvas púrpuras representam os contornos de decisão de Bayes

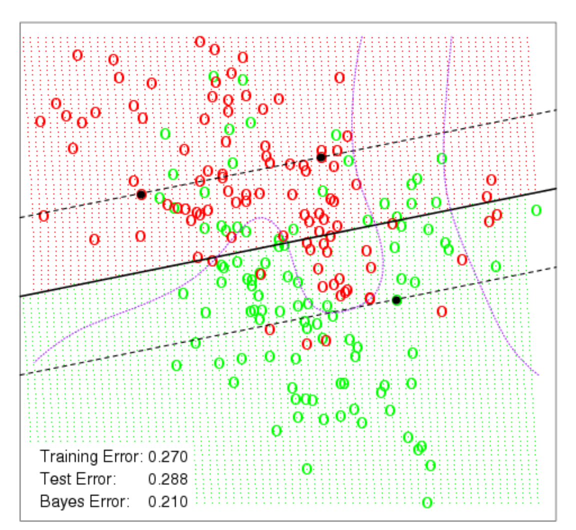


SVM

Exemplo constante maior

- Usando uma constante maior permite uma margem mais larga e cria um classificador diferente.
- Note, contudo, que o contorno de decisão mantém-se linear.

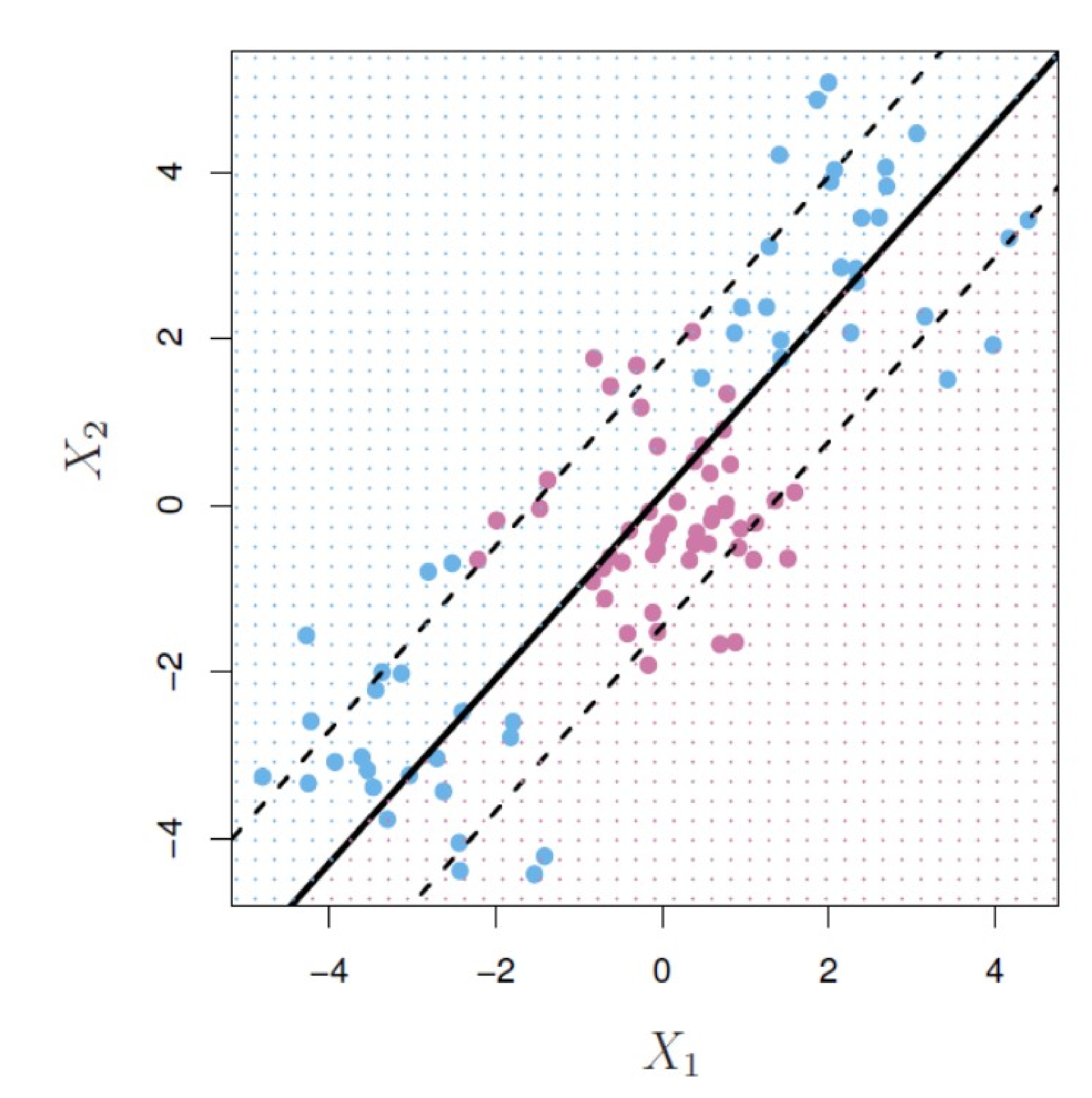




Classificador por “Máquina de Vetor de Suporte”

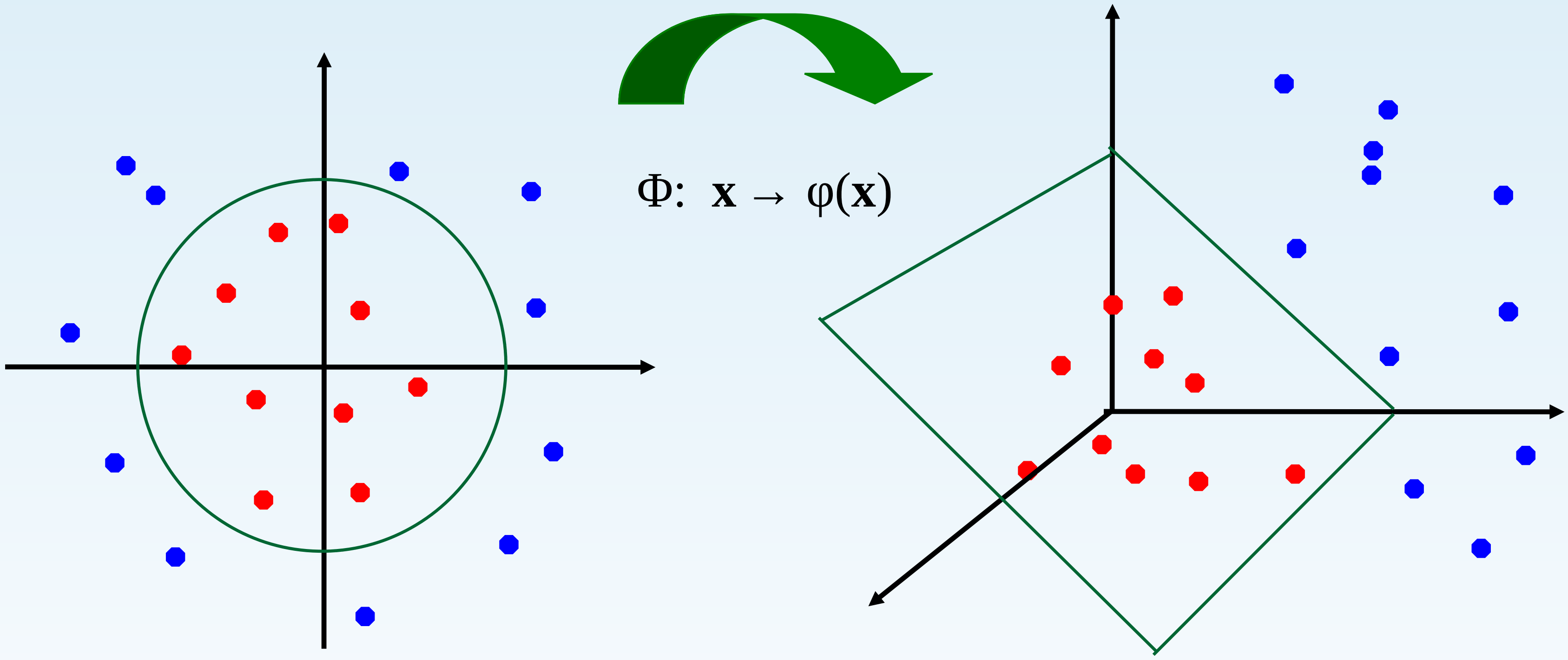
Classificador por “Máquina de Vetor de Suporte”

- Às vezes, independente da escolha de C, uma separação linear não existirá.
- O que fazer?



SVM Não-linear

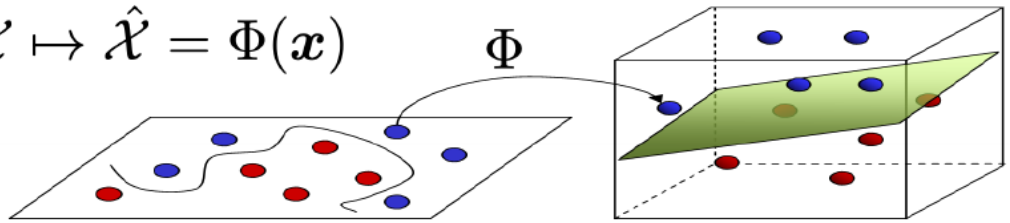
- Ideia geral: o espaço de entrada original pode ser mapeado para algum espaço de atributos de maior dimensão onde o conjunto de treinamento seja separável:



SVM Não-linear

Mapping into a New Feature Space

$\Phi : \mathcal{X} \mapsto \hat{\mathcal{X}} = \Phi(\boldsymbol{x})$



- For example, with $\boldsymbol{x} \in \mathbb{R}^2$, let $\Phi([x_{i1}, x_{i2}]) = [x_{i1}, x_{i2}, x_{i1}x_{i2}, x_{i1}^2, x_{i2}^2]$
- Rather than run SVM on $\boldsymbol{x}_i$, run it on $\Phi(\boldsymbol{x}_i)$
  - Finds a non-linear separator in the input space

This is just a basis expansion

What if $\Phi(\boldsymbol{x}_i)$ is really big?

- Use kernels to compute it implicitly!

imagem/slide: E.Eaton

Classificador por “Máquina de Vetor de Suporte”

- O classificador de vetor de suporte permite, contudo, somente contorno linear de decisão.

Classificador por “Máquina de Vetor de Suporte”

- O classificador de vetor de suporte permite, contudo, somente contorno linear de decisão.
- Pode-se estender regressão linear para regressão não-linear usando funções de base $b_i(X_i)$ i.e.

$$Y_i = \beta_0 + \beta_1 b_1(X_i) + \beta_2 b_2(X_i) + \cdots + \beta_p b_p(X_i) + \varepsilon_i$$

"Kernel trick"

- O classificador linear suporta-se no produto interno entre vetores $K(\mathrm{x}_i,\mathrm{x}_j)=\mathrm{x}_i^T\mathrm{x}_j$
- Se todo ponto de dados é transformado em um espaço de maior dimensão via alguma transformação Φ: $\mathrm{x} \rightarrow \varphi(\mathrm{x})$, o produto interno torna-se:

$$K(\mathrm{x}_i,\mathrm{x}_j)= \varphi(\mathrm{x}_i)^T\varphi(\mathrm{x}_j)$$

- Uma *função de kernel* é alguma função que corresponda ao produto interno em algum espaço de atributos expandido.
- Exemplo:

vetores bidimensionais $\mathrm{x}=[x_1\ \ x_2]$; Seja $K(\mathrm{x}_i,\mathrm{x}_j)=(1 + \mathrm{x}_i^T\mathrm{x}_j)^2$,

Necessário mostrar que $K(\mathrm{x}_i,\mathrm{x}_j)= \varphi(\mathrm{x}_i)^T\varphi(\mathrm{x}_j)$:

$$\begin{aligned} K(\mathrm{x}_i,\mathrm{x}_j)&=(1 + \mathrm{x}_i^T\mathrm{x}_j)^2, \\ &= 1+ x_{i1}^2x_{j1}^2 + 2\, x_{i1}x_{j1}\, x_{i2}x_{j2}+ x_{i2}^2x_{j2}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} \\ &= [1\ \ x_{i1}^2\ \ \sqrt{2}\, x_{i1}x_{i2}\ \ x_{i2}^2\ \ \sqrt{2}x_{i1}\ \ \sqrt{2}x_{i2}]^T [1\ \ x_{j1}^2\ \ \sqrt{2}\, x_{j1}x_{j2}\ \ x_{j2}^2\ \ \sqrt{2}x_{j1}\ \ \sqrt{2}x_{j2}] \\ &= \varphi(\mathrm{x}_i)^T\varphi(\mathrm{x}_j), \quad \text{onde } \varphi(\mathrm{x}) = [1\ \ x_1^2\ \ \sqrt{2}\, x_1x_2\ \ x_2^2\ \ \sqrt{2}x_1\ \ \sqrt{2}x_2] \end{aligned}$$

Classificador por “Máquina de Vetor de Suporte”: Expansão de Atributos

- Aumentar o espaço de atributos (características) incluindo transformações: $X_1^2,\ X_1^3,\ X_1X_2,\ X_1X_2^2, \ldots$
- Indo de um espaço com *p* dimensões a *M* > *p*
- Ajustar um classificador SVM no espaço aumentado
- Isso resultará em contornos não-lineares de decisão no espaço original
- Exemplo: $(X_1,\ X_2,\ X_1^2,\ X_2^2,\ X_1X_2)$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 = 0$$

Abordagem básica SVM não-linear

- Conceitualmente, fazemos uma abordagem similar com o classificador vetor de suporte.
- O classificador vetor de suporte encontra o hiperplano ótimo no espaço gerado por X_1, X_2,…, X_p.

Abordagem básica SVM não-linear

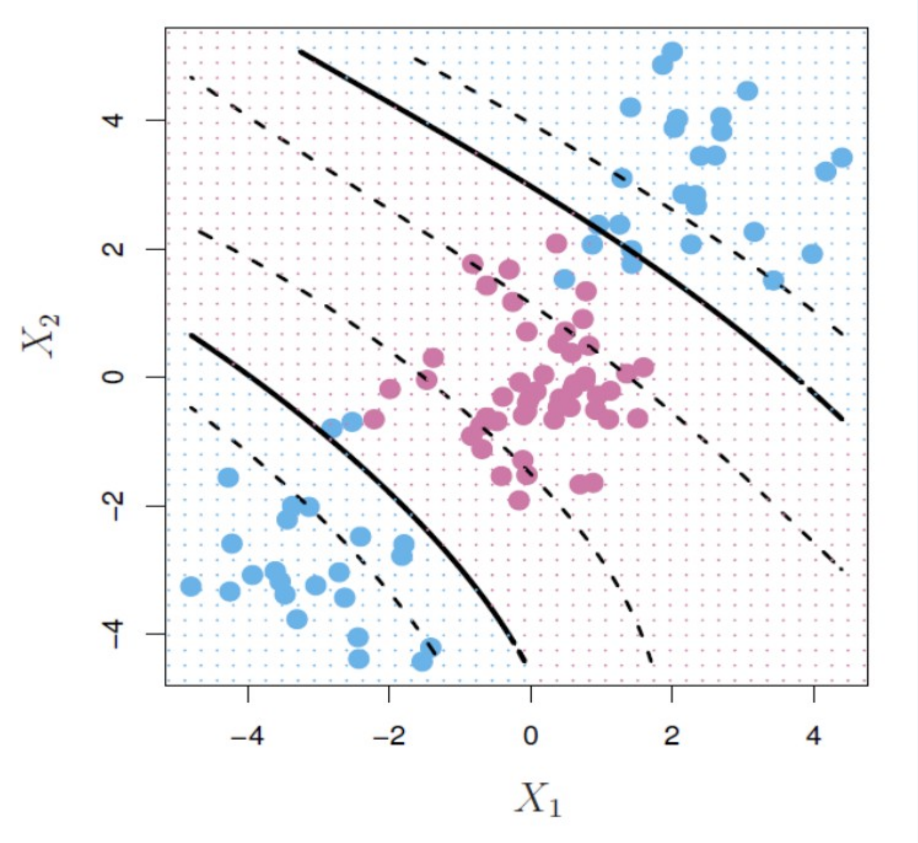
- Conceitualmente, fazemos uma abordagem similar com o classificador vetor de suporte.
- O classificador vetor de suporte encontra o hiperplano ótimo no espaço gerado por $X_1, X_2, \ldots, X_p$.
- Podemos criar transformações (ou uma base) $b_1(x), b_2(x), \ldots, b_M(x)$ e encontrar o hiperplano ótimo no espaço gerado por $b_1(\mathbf{X}), b_2(\mathbf{X}), \ldots, b_M(\mathbf{X})$.

Abordagem básica SVM não-linear

- Conceitualmente, fazemos uma abordagem similar com o classificador vetor de suporte.
- O classificador vetor de suporte encontra o hiperplano ótimo no espaço gerado por $X_1, X_2, \ldots, X_p$.
- Podemos criar transformações (ou uma base) $b_1(x)$, $b_2(x)$, …, $b_M(x)$ e encontrar o hiperplano óptimo no espaço gerado por $b_1(\mathbf{X})$, $b_2(\mathbf{X})$, …, $b_M(\mathbf{X})$.
- Essa abordagem produz um plano linear no espaço transformado mas um contorno de decisão não-linear no espaço original.
- Isso é chamado de Classificador por Máquina de Vetor de Suporte.

Exemplo

Polinômios cúbicos



$$\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1^2+\beta_4X_2^2+\beta_5X_1X_2+\beta_6X_1^3+\beta_7X_2^3+\beta_8X_1X_2^2+\beta_9X_1^2X_2=0$$

Mas, na prática

- Conceitualmente a abordagem básica é como SVM funciona realmente, mas não escolhemos $b_1(x), b_2(x), \ldots, b_M(x)$.
- Ao invés, escolhemos algo chamado de função Kernel (ou radial) que toma o lugar da base.
- Funções comuns "kernel" incluem
  - Linear
  - Polinomial
  - Base Radial
  - Sigmóide

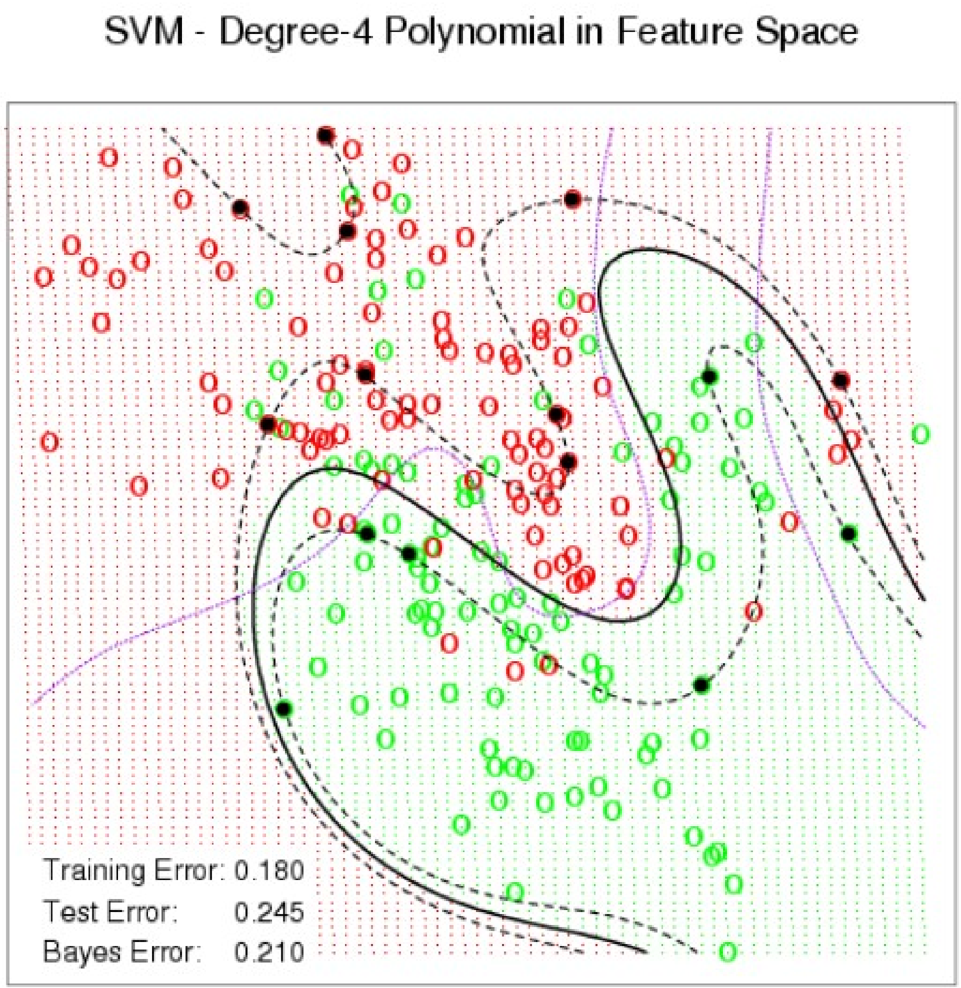
Mas, na prática

- Ao invés, escolhemos algo chamado de função Kernel (ou radial) que toma o lugar da base.
- Funções comuns "kernel" incluem
  - Linear $K(\mathbf{x_i},\mathbf{x_j}) = \mathbf{x_i}^T\mathbf{x_j}$
  - Polinomial $K(\mathbf{x_i},\mathbf{x_j}) = (1 + \mathbf{x_i}^T\mathbf{x_j})^p$
  - Base Radial $K(\mathbf{x_i},\mathbf{x_j}) = \exp\left(-\dfrac{\|\mathbf{x_i} - \mathbf{x_j}\|^2}{2\sigma^2}\right)$
  - Sigmóide $K(\mathbf{x_i},\mathbf{x_j}) = \tanh(\beta_0\mathbf{x_i}^T\mathbf{x_j} + \beta_1)$

Exemplo

Exemplo com kernel polinomial

- Usando um kernel polinomial permitimos ao SVM produzir um contorno de decisão não-linear.
- Note que a taxa de erro de teste é bem menor.

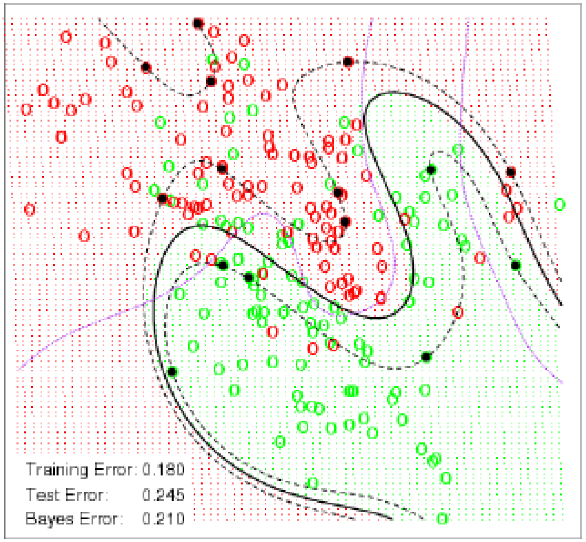


Exemplo

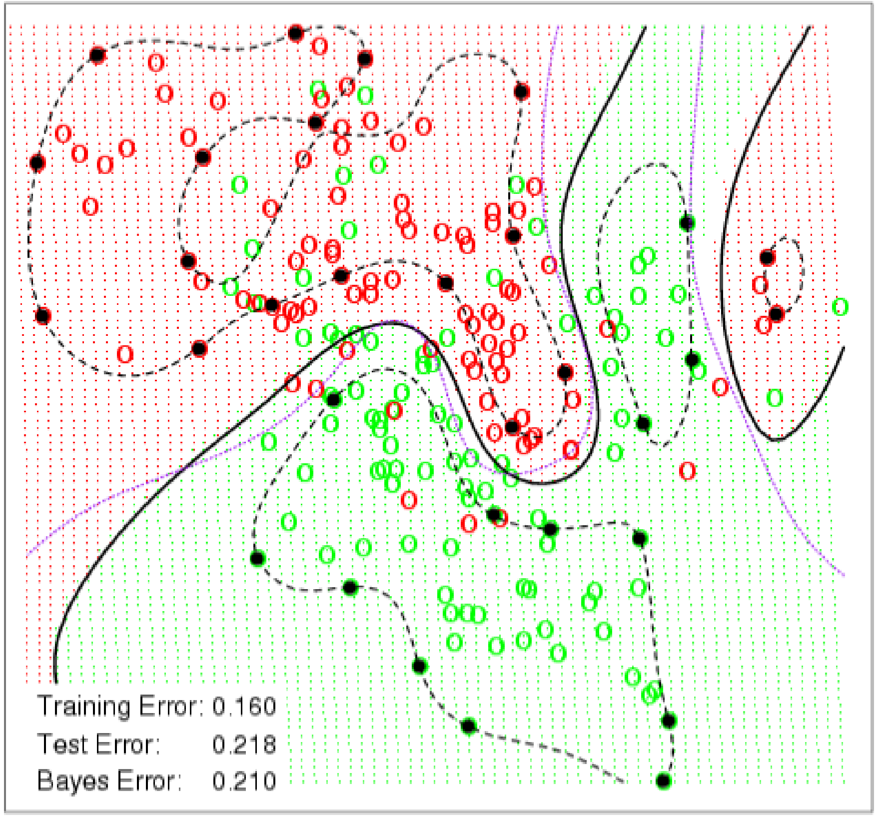
Exemplo, base radial

- Usando um Kernel de Base Radial consegue-se uma taxa de erro ainda menor.

SVM - Degree-4 Polynomial in Feature Space



SVM - Radial Kernel in Feature Space



SVM para mais de 2 classes?

OVA One versus All. Fit K different 2-class SVM classifiers $\hat{f}_k(x)$, $k = 1, \ldots, K$; each class versus the rest. Classify x^* to the class for which $\hat{f}_k(x^*)$ is largest.

SVM para mais de 2 classes?

OVA One versus All. Fit K different 2-class SVM classifiers $\hat{f}_k(x)$, $k = 1, \ldots, K$; each class versus the rest. Classify x^* to the class for which $\hat{f}_k(x^*)$ is largest.

OVO One versus One. Fit all $\binom{K}{2}$ pairwise classifiers $\hat{f}_{k\ell}(x)$. Classify x^* to the class that wins the most pairwise competitions.

Exemplo de Aplicação

Application: Pedestrian detection in Computer Vision

Objective: detect (localize) standing humans in an image

- cf face detection with a sliding window classifier

- reduces object detection to binary classification
- does an image window contain a person or not?

Method: the HOG detector

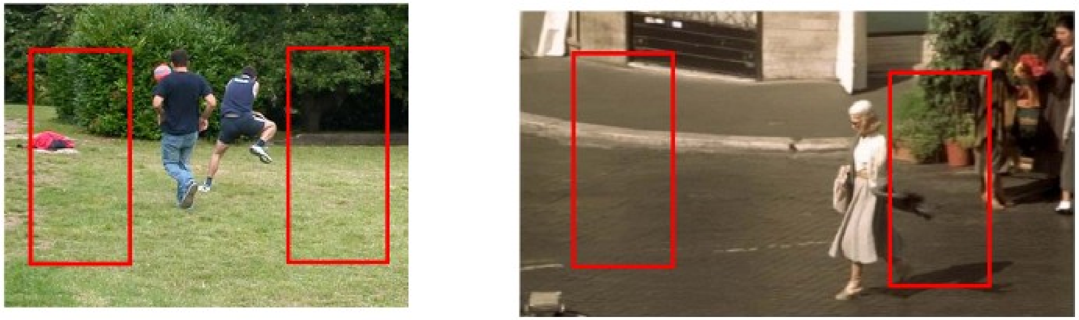
(Dalal & Triggs, 2005)

Exemplo de Aplicação

Training data and features

- Positive data – 1208 positive window examples

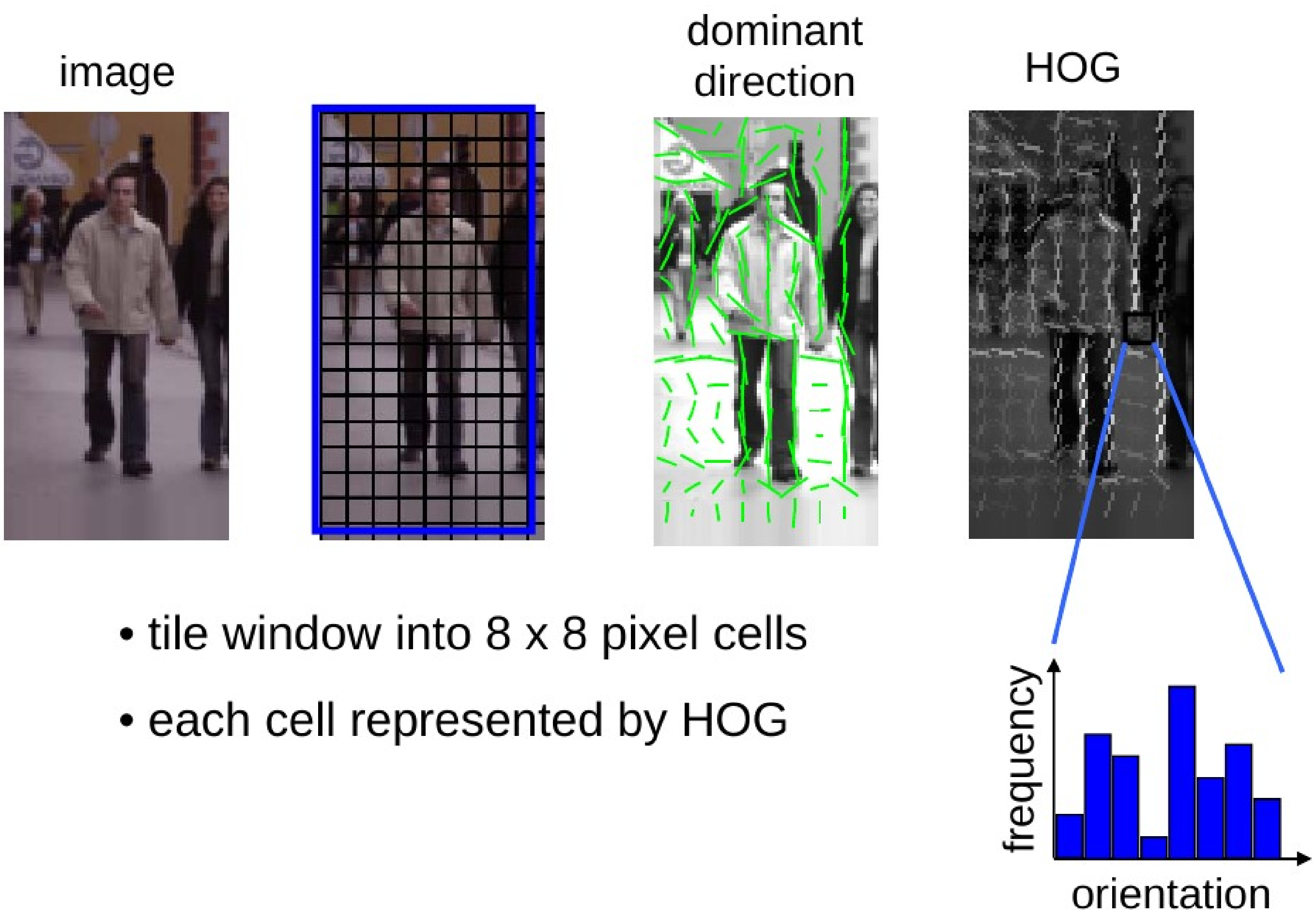
- Negative data – 1218 negative window examples (initially)



(Dalal & Triggs, 2005)

Exemplo de Aplicação

Feature: histogram of oriented gradients (HOG)

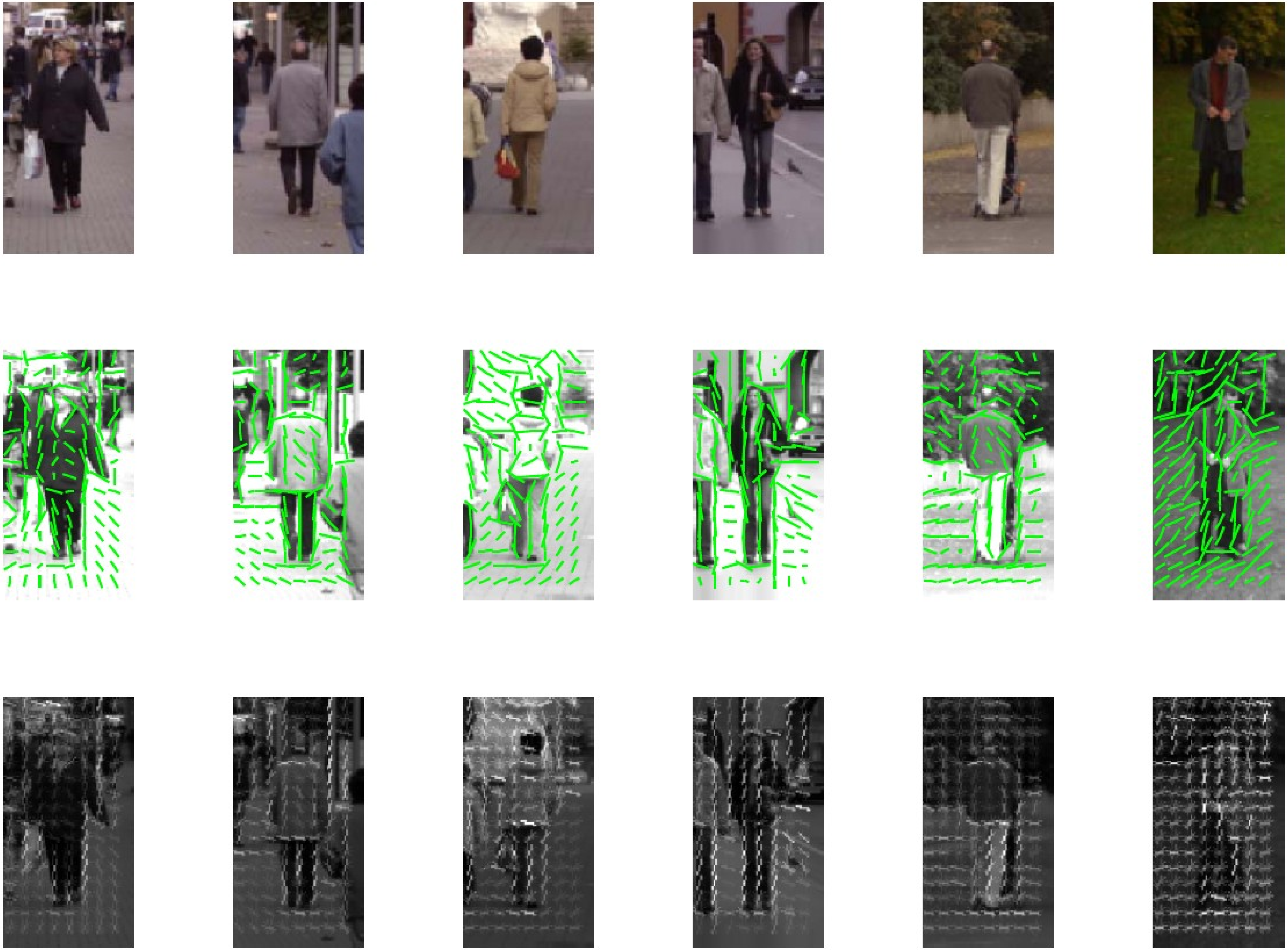


- tile window into 8 x 8 pixel cells
- each cell represented by HOG

Feature vector dimension = 16 x 8 (for tiling) x 8 (orientations) = 1024

(Dalal & Triggs, 2005)

Exemplo de Aplicação



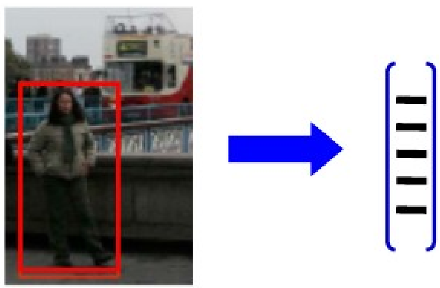
(Dalal & Triggs, 2005)

Exemplo de Aplicação

Algorithm

Training (Learning)

- Represent each example window by a HOG feature vector



$$\mathbf{x}_i \in \mathbb{R}^d, \text{ with } d = 1024$$

- Train a SVM classifier

Testing (Detection)

- Sliding window classifier

$$f(x) = \mathbf{w}^\top \mathbf{x} + b$$

(Dalal & Triggs, 2005)

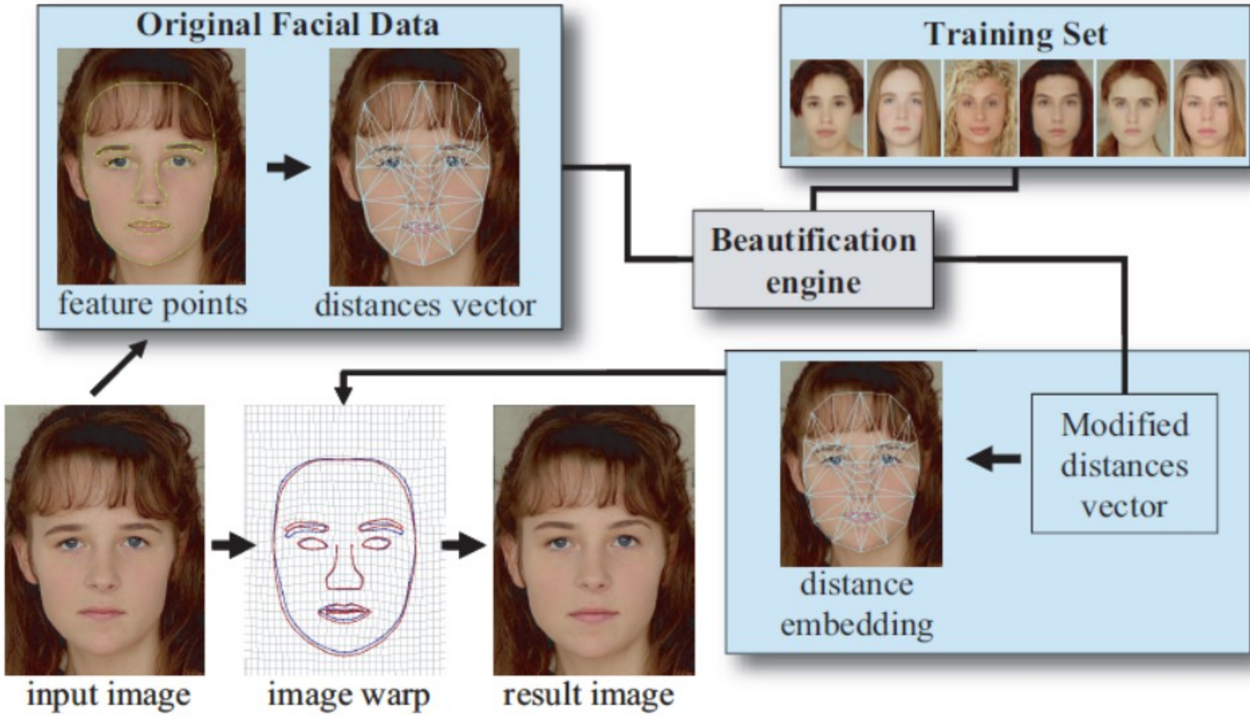
Exemplo de Aplicação

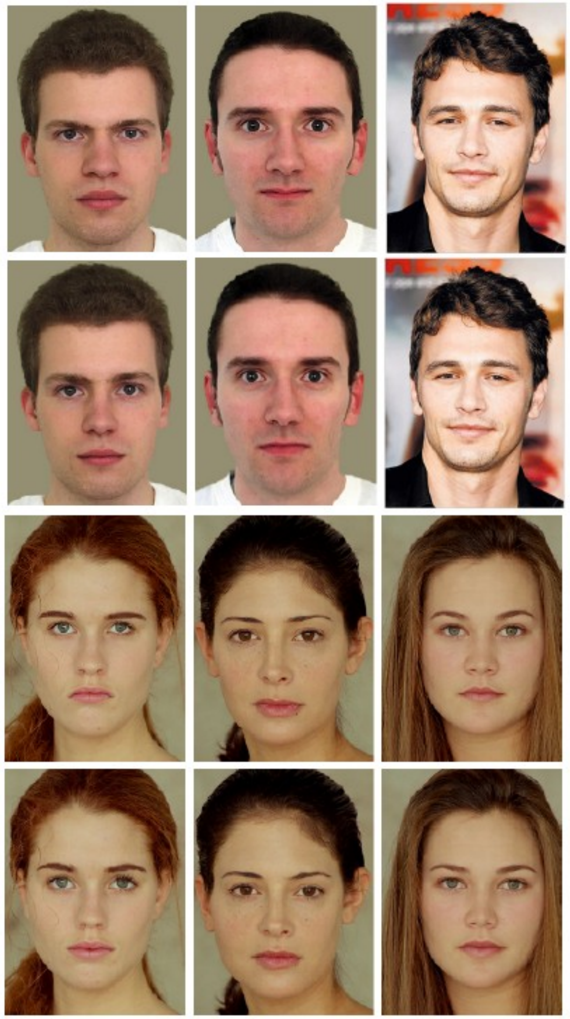
Dalal and Triggs, CVPR 2005

Exemplo de Aplicação

Application: Automatic Photo Retouching

[Leyvand et al., 2008]





Sugestão de Leitura

- **Ler o Capítulo 9 do livro "*James, Witten, Hastie & Tibshirani, Introduction to Statistical Learning with applications in R, Springer, 2017*."**

- **Capítulo 19 do livro do Russell & Norvig, *Artificial Intelligence: a modern approach*, 4th ed., Pearson, 2020.**

Referências Bibliográficas

- Alpaydin, E. *Introduction to Machine Learning*. MIT Press, 2010.
- Bishop, C. *Pattern Recognition and Machine Learning.* Springer, 2006.
- James, G.; Witten, D.; Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning with applications in R,* Springer, 2017.
- Mitchell, T. *Machine Learning.* McGraw Hill, 1997.