Inteligência Artificial – ACH2016
Aula21 – Support Vector Machines

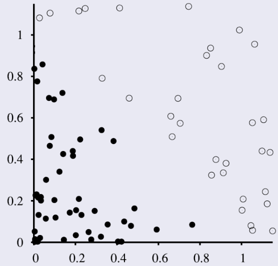
Norton Trevisan Roman
(norton@usp.br)

27 de maio de 2019

Support Vector Machine – SVM

Margem Máxima

- Considere o conjunto linearmente separável ao lado

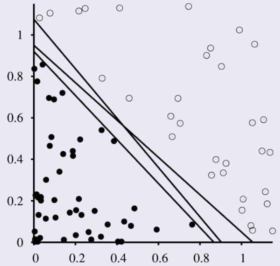


Fonte: AIMA. R&N

Support Vector Machine – SVM

Margem Máxima

- Considere o conjunto linearmente separável ao lado
  - Temos várias possibilidades de separação

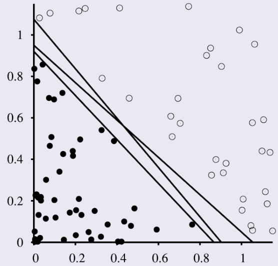


Fonte: AIMA. R&N

Support Vector Machine – SVM

Margem Máxima

- Considere o conjunto linearmente separável ao lado
  - Temos várias possibilidades de separação
  - Qual seria a melhor?

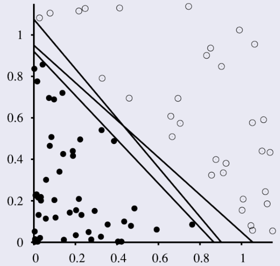


Fonte: AIMA. R&N

Support Vector Machine – SVM

Margem Máxima

- Considere o conjunto linearmente separável ao lado
  - Temos várias possibilidades de separação
  - Qual seria a melhor?



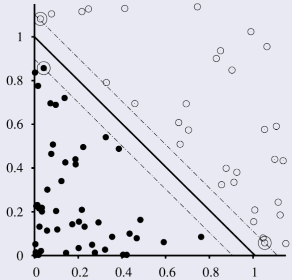
Fonte: AIMA. R&N

- Cada linha que separa os dados é um **hiperplano de separação**
  - Será nosso limite de decisão → tudo de um lado pertence a uma classe, e tudo do outro pertence a outra

Support Vector Machine – SVM

Margem Máxima

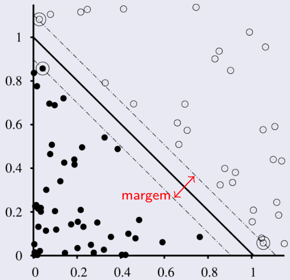
- Gostaríamos de escolher o separador que estivesse o mais longe possível dos exemplos
  - Acomodando, assim, possíveis erros de classificação



Fonte: AIMA. R&N

Margem Máxima

- Gostaríamos de escolher o separador que estivesse o mais longe possível dos exemplos
  - Acomodando, assim, possíveis erros de classificação
  - Esse seria então um **separador de margem máxima**

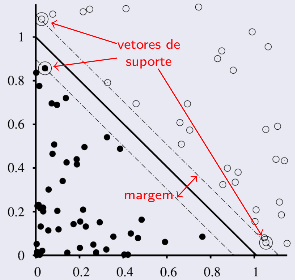


Fonte: AIMA. R&N

Support Vector Machine – SVM

Margem Máxima

- Gostaríamos de escolher o separador que estivesse o mais longe possível dos exemplos
  - Acomodando, assim, possíveis erros de classificação
  - Esse seria então um **separador de margem máxima**



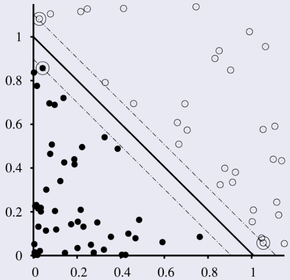
Fonte: AIMA. R&N

- Os pontos mais próximos do separador são seus **vetores de suporte**
  - Queremos então maximizar a distância entre esse hiperplano e seus vetores de suporte

Support Vector Machine – SVM

Classificador SVM

- Baseia-se na ideia principal do separador de margem máxima
  - Quanto mais longe um ponto está do limite de decisão, mais confiantes estamos sobre a predição feita

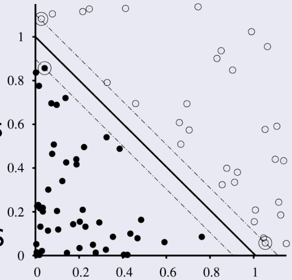


Fonte: AIMA. R&N

Support Vector Machine – SVM

Classificador SVM

- Baseia-se na ideia principal do separador de margem máxima
  - Quanto mais longe um ponto está do limite de decisão, mais confiantes estamos sobre a predição feita
- Difere dos demais classificadores em que retorna $+1$ ou -1 em sua versão binária
  - Isso acaba facilitando os cálculos...



Fonte: AIMA. R&N

Support Vector Machine – SVM

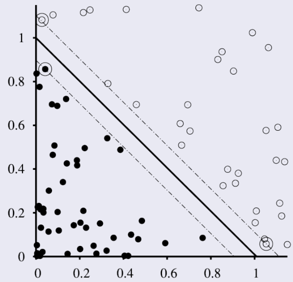
SVM linear com margens rígidas

- Define fronteiras lineares a partir de dados linearmente separáveis
  - Não permite pontos nessa fronteira

Support Vector Machine – SVM

SVM linear com margens rígidas

- Define fronteiras lineares a partir de dados linearmente separáveis
  - Não permite pontos nessa fronteira

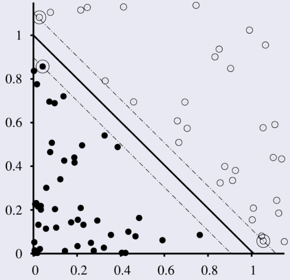


Fonte: AIMA. R&N

Support Vector Machine – SVM

SVM linear com margens rígidas

- Define fronteiras lineares a partir de dados linearmente separáveis
  - Não permite pontos nessa fronteira



Fonte: AIMA. R&N

- Um classificador linear é aquele que separa os dados com um hiperplano do tipo $f(\vec{x}) = \vec{\omega} \cdot \vec{x} + b$
  - A equação separa o espaço de dados X em duas regiões, $\vec{\omega} \cdot \vec{x} + b \geq 0$ e $\vec{\omega} \cdot \vec{x} + b < 0$

SVM Linear com Margens Rígidas

Hiperplano Canônico

- Usamos então uma função sinal $g(\vec{x}) = sgn(f(\vec{x}))$ para classificar um ponto $\vec{x}$:

$$g(\vec{x}) = sgn(f(\vec{x})) = \begin{cases} +1, & \text{se } \vec{\omega} \cdot \vec{x} + b \geq 0 \\ -1, & \text{se } \vec{\omega} \cdot \vec{x} + b < 0 \end{cases}$$

SVM Linear com Margens Rígidas

Hiperplano Canônico

- Usamos então uma função sinal $g(\vec{x}) = sgn(f(\vec{x}))$ para classificar um ponto $\vec{x}$:

$$g(\vec{x}) = sgn(f(\vec{x})) = \begin{cases} +1, & \text{se } \vec{\omega} \cdot \vec{x} + b \geq 0 \\ -1, & \text{se } \vec{\omega} \cdot \vec{x} + b < 0 \end{cases}$$

- Hiperplano canônico:
  - Aquele em que $\vec{\omega}$ e b são escolhidos de forma que os exemplos mais próximos do hiperplano satisfaçam a equação $|\vec{\omega} \cdot \vec{x} + b| = 1$
  - Define a margem

SVM Linear com Margens Rígidas

Hiperplano Canônico

- Então, temos 2 hiperplanos nas bordas da margem:

$$\begin{cases} \vec{\omega} \cdot \vec{x} + b \geq +1, & \text{se } y_i = +1 \\ \vec{\omega} \cdot \vec{x} + b \leq -1, & \text{se } y_i = -1 \end{cases}$$

onde $y_i \in Y$, $Y = \{-1, +1\}$ é o rótulo de x_i

SVM Linear com Margens Rígidas

Hiperplano Canônico

- Então, temos 2 hiperplanos nas bordas da margem:

$$\begin{cases} \vec{\omega} \cdot \vec{x} + b \geq +1, & \text{se } y_i = +1 \\ \vec{\omega} \cdot \vec{x} + b \leq -1, & \text{se } y_i = -1 \end{cases}$$

  onde $y_i \in Y$, $Y = \{-1, +1\}$ é o rótulo de x_i

- E aqui vemos a vantagem de se definir as classes como $Y = \{-1, +1\}$
  - Podemos resumir a expressão acima em uma única equação:

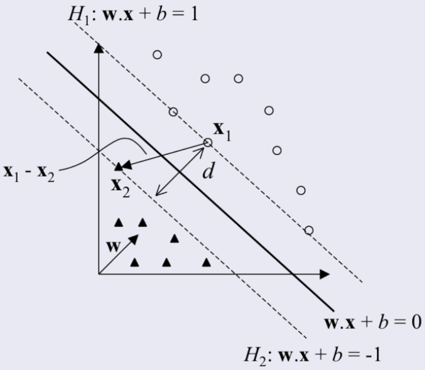
  $y_i(\vec{\omega} \cdot \vec{x} + b) - 1 \geq 0, \forall (\vec{x}_i, y_i) \in T$

  (onde T é o conjunto de treino)

SVM Linear com Margens Rígidas

Distância entre os Hiperplanos

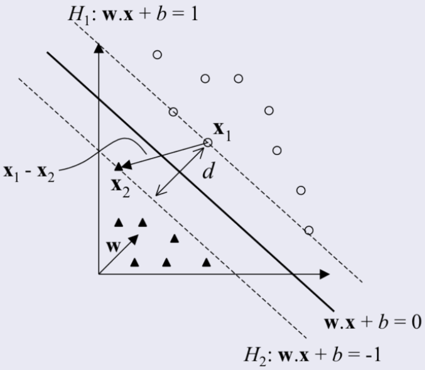
- Queremos, no entanto, maximizar a margem
  - Maximizar a distância d entre os hiperplanos H_1 e H_2



Fonte: [3]

Distância entre os Hiperplanos

- Queremos, no entanto, maximizar a margem
  - Maximizar a distância d entre os hiperplanos H_1 e H_2
- Projetamos então o vetor $\vec{x}_1 - \vec{x}_2$ na direção de $\vec{\omega}$
  - Onde $\vec{x}_1 \in H_1$ e $\vec{x}_2 \in H_2$,



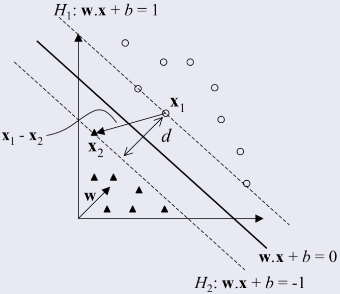
Fonte: [3]

SVM Linear com Margens Rígidas

Distância entre os Hiperplanos

- E

$$\vec{d} = (\vec{x}_1 - \vec{x}_2)\left(\frac{\vec{\omega}}{\|\vec{\omega}\|} \cdot \frac{(\vec{x}_1 - \vec{x}_2)}{\|\vec{x}_1 - \vec{x}_2\|}\right)$$



Fonte: [3]

SVM Linear com Margens Rígidas

Distância entre os Hiperplanos

- E

$$\vec{d} = (\vec{x}_1 - \vec{x}_2)\left(\frac{\vec{\omega}}{\|\vec{\omega}\|} \cdot \frac{(\vec{x}_1 - \vec{x}_2)}{\|\vec{x}_1 - \vec{x}_2\|}\right)$$

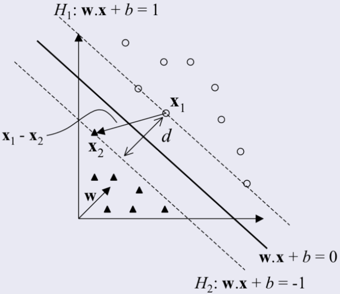
- Uma vez que

$\vec{\omega} \cdot \vec{x}_1 + b = +1$ e $\vec{\omega} \cdot \vec{x}_2 + b = -1$

- Então

$$\vec{\omega} \cdot (\vec{x}_1 - \vec{x}_2) = \vec{\omega} \cdot \vec{x}_1 - \vec{\omega} \cdot \vec{x}_2$$

$$= (1 - b) - (-1 - b) = 2$$

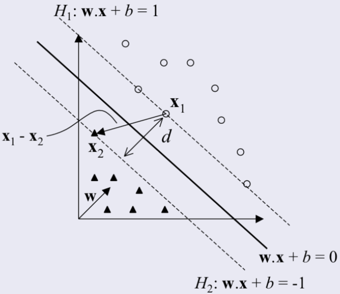


Fonte: [3]

SVM Linear com Margens Rígidas

Distância entre os Hiperplanos

- E $\vec{d} = \dfrac{2(\vec{x}_1 - \vec{x}_2)}{\|\vec{\omega}\| \|\vec{x}_1 - \vec{x}_2\|}$



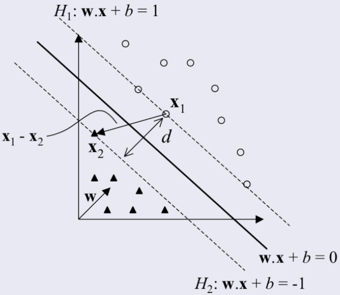
Fonte: [3]

SVM Linear com Margens Rígidas

Distância entre os Hiperplanos

- E $\vec{d} = \dfrac{2(\vec{x}_1 - \vec{x}_2)}{\|\vec{\omega}\|\|\vec{x}_1 - \vec{x}_2\|}$
- Assim, $d = \|\vec{d}\| = \dfrac{2}{\|\vec{\omega}\|}$

  $d' = \frac{1}{\|\vec{\omega}\|}$ é então a distância mínima entre o hiperplano separador e os dados de treinamento



Fonte: [3]

SVM Linear com Margens Rígidas

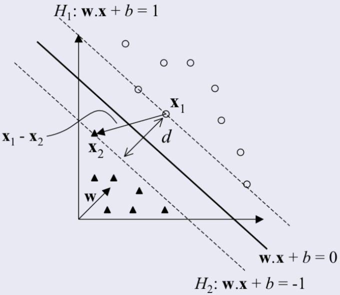
Distância entre os Hiperplanos

- E $\vec{d} = \dfrac{2(\vec{x}_1 - \vec{x}_2)}{\|\vec{\omega}\| \|\vec{x}_1 - \vec{x}_2\|}$

- Assim, $d = \|\vec{d}\| = \dfrac{2}{\|\vec{\omega}\|}$

  $d' = \frac{1}{\|\vec{\omega}\|}$ é então a distância mínima entre o hiperplano separador e os dados de treinamento

- Queremos maximizar d'



Fonte: [3]

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Maximizar $d' = 1/\|\vec{\omega}\|$ corresponde ao problema de otimização:

$$f(\vec{x}) = \min_{\vec{\omega},b} \left(\frac{1}{2} \|\vec{\omega}\|^2 \right)$$

Restrição: $y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 \geq 0$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Maximizar $d' = 1/\|\vec{\omega}\|$ corresponde ao problema de otimização:

$$f(\vec{x}) = \min_{\vec{\omega},b} \left(\frac{1}{2} \|\vec{\omega}\|^2 \right)$$

Restrição: $y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 \geq 0$

A restrição garante que não haja dados de treino entre as margens de separação das classes

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Maximizar $d' = 1/\|\vec{\omega}\|$ corresponde ao problema de otimização:

  $$f(\vec{x}) = \min_{\vec{\omega},b} \left(\frac{1}{2} \|\vec{\omega}\|^2 \right)$$

  Restrição: $y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 \geq 0$

- Por que isso?
  - Porque os mesmos $\vec{\omega}$ e b que resolvem um problema também resolve o outro

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Maximizar $d' = 1/\|\vec{\omega}\|$ corresponde ao problema de otimização:

$$f(\vec{x}) = \min_{\vec{\omega},b} \left(\frac{1}{2} \|\vec{\omega}\|^2 \right)$$

Restrição: $y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 \geq 0$

Fonte: https://brainsnorts.files.wordpress.com/2014/05/calvin376_2.jpg

- Por que isso?
  - Porque os mesmos $\vec{\omega}$ e b que resolvem um problema também resolve o outro
  - E alguém já quebrou a cabeça resolvendo um deles

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- A solução desse problema de otimização passa pela introdução de uma Lagrangiana

$$L(\vec{\omega}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{\omega}\|^2 - \sum_{i=1}^{n} \alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1)$$

Onde α_i são os chamados multiplicadores de Lagrange

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- A solução desse problema de otimização passa pela introdução de uma Lagrangiana

$$L(\vec{\omega}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{\omega}\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1)$$

  Onde α_i são os chamados multiplicadores de Lagrange

- $L(\vec{\omega}, b, \vec{\alpha})$ deve então ser minimizada
  - Para isso, maximizamos α_i e minimizamos $\vec{\omega}$ e b
  - Para $\vec{\omega}$ e b, fazemos $\frac{\partial L}{\partial b} = 0$ e $\frac{\partial L}{\partial \vec{\omega}} = 0$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- O que nos leva ao resultado

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \text{ e } \vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- O que nos leva ao resultado

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \text{ e } \vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$$

- Para α_i, substituímos esse resultado na Lagrangeana e maximizamos
  - Queremos então $\max\limits_{\vec{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$

Com as restrições $\begin{cases} \alpha_i \geq 0, \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- O que nos leva ao resultado

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \text{ e } \vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$$

- Para α_i, substituímos esse resultado na Lagrangeana e maximizamos
  - Queremos então $\max_{\vec{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$

Com as restrições $\begin{cases} \alpha_i \geq 0, \quad i = 1, \ldots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$

Formulação conhecida como **forma dual** do problema

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Assim, encontrada a solução $\vec{\alpha}$ da forma dual, usamos $\vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$ para achar a solução $\vec{\omega}$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Assim, encontrada a solução $\vec{\alpha}$ da forma dual, usamos $\vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$ para achar a solução $\vec{\omega}$
- E b?
  - Obtido de α e das condições de Kühn-Tucker (teoria de otimização com restrições)
  - Para esse problema, temos as restrições $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$, $i = 1, \ldots, n$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Vejamos a restrição $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$
- Temos que $\alpha_i \neq 0$ apenas para pontos sobre H_1 e H_2
  - Os exemplos mais próximos do hiperplano separador

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Vejamos a restrição $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$
- Temos que $\alpha_i \neq 0$ apenas para pontos sobre H_1 e H_2
  - Os exemplos mais próximos do hiperplano separador
- Para os demais deve ser 0. Por que?

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Vejamos a restrição $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$
- Temos que $\alpha_i \neq 0$ apenas para pontos sobre H_1 e H_2
  - Os exemplos mais próximos do hiperplano separador
- Para os demais deve ser 0. Por que?
  - Se $\vec{x} \notin H_1$ ou H_2, então $y_i(\vec{\omega} \cdot \vec{x} + b) - 1 > 0$
  - Não está nem na margem, nem no plano

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Vejamos a restrição $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$
- Temos que $\alpha_i \neq 0$ apenas para pontos sobre H_1 e H_2
  - Os exemplos mais próximos do hiperplano separador
- Para os demais deve ser 0. Por que?
  - Se $\vec{x} \notin H_1$ ou H_2, então $y_i(\vec{\omega} \cdot \vec{x} + b) - 1 > 0$
  - Não está nem na margem, nem no plano
  - A única forma de $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1)$ ser 0 é se $\alpha_i = 0$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Os exemplos em que $\alpha_i > 0$ são os vetores de suporte V para o hiperplano separador
  - Apenas eles participarão da determinação da equação desse hiperplano

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Os exemplos em que $\alpha_i > 0$ são os vetores de suporte V para o hiperplano separador
  - Apenas eles participarão da determinação da equação desse hiperplano
- Mas e b? Calculado de $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$

$$\begin{aligned} y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 &= 0 \\ \vec{\omega} \cdot \vec{x}_i + b &= 1/y_i \\ b &= 1/y_i - \vec{\omega} \cdot \vec{x}_i \end{aligned}$$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Os exemplos em que $\alpha_i > 0$ são os vetores de suporte V para o hiperplano separador
  - Apenas eles participarão da determinação da equação desse hiperplano
- Mas e b? Calculado de $\alpha_i(y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1) = 0$

$$
\begin{aligned}
y_i(\vec{\omega} \cdot \vec{x}_i + b) - 1 &= 0 \\
\vec{\omega} \cdot \vec{x}_i + b &= 1/y_i \\
b &= 1/y_i - \vec{\omega} \cdot \vec{x}_i
\end{aligned}
$$

  - Isso, contudo considerando apenas um vetor de suporte

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Como temos n_V vetores de suporte, b será a média dentre eles
$$b = \frac{1}{n_V} \sum_{x_j \in V} \frac{1}{y_j} - \vec{\omega} \cdot \vec{x}_j$$

SVM Linear com Margens Rígidas

Maximizando a distância entre os hiperplanos

- Como temos n_V vetores de suporte, b será a média dentre eles

$$b = \frac{1}{n_V} \sum_{x_j \in V} \frac{1}{y_j} - \vec{\omega} \cdot \vec{x}_j$$

- E como $\vec{\omega} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i$, então

$$b = \frac{1}{n_V} \sum_{x_j \in V} \left(\frac{1}{y_j} - \sum_{x_i \in V}^{n} \alpha_i y_i \vec{x}_i \cdot \vec{x}_j \right)$$

SVM Linear com Margens Rígidas

RESUME ISSO POR FAVOR!!!!

Fonte: https://i.imgflip.com/pqcjo.jpg?a432792

SVM Linear com Margens Rígidas

O classificador SVM

- Dado um ponto $\vec{x}$, sua classificação será dada por

$$g(\vec{x}) = sgn(f(\vec{x})) = sgn(\vec{\omega} \cdot \vec{x} + b) = sgn\left(\sum_{\vec{x_i} \in V} y_i \alpha_i \vec{x_i} \cdot \vec{x} + b\right)$$

(com α, $\vec{\omega}$ e b calculados como mostrado)

SVM Linear com Margens Rígidas

O classificador SVM

- Dado um ponto $\vec{x}$, sua classificação será dada por

$$g(\vec{x}) = sgn(f(\vec{x})) = sgn(\vec{\omega} \cdot \vec{x} + b) = sgn\left(\sum_{\vec{x_i} \in V} y_i \alpha_i \vec{x_i} \cdot \vec{x} + b\right)$$

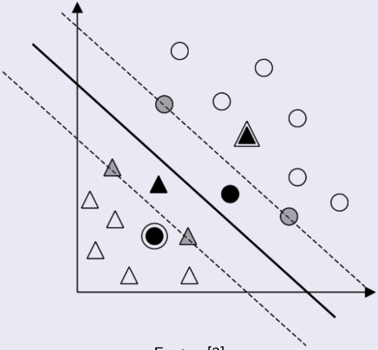
  (com α, $\vec{\omega}$ e b calculados como mostrado)

- Esse é o classificador SVM
  - Representando o hiperplano que separa os dados com maior margem

SVM Linear com Margens Suaves

Dados com ruído

- Em algumas situações, mesmo sendo linearmente separáveis, os dados apresentam ruídos



Fonte: [3]

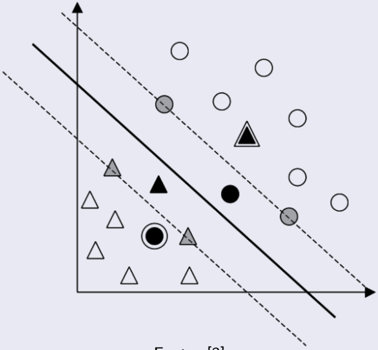
SVM Linear com Margens Suaves

Dados com ruído

- Em algumas situações, mesmo sendo linearmente separáveis, os dados apresentam ruídos
- Para esses casos, relaxamos as restrições do SVM

$$y_i(\vec{\omega} \cdot \vec{x}_i + b) \geq 1 - \xi_i$$

onde $\xi_i \geq 0$ é uma variável de folga

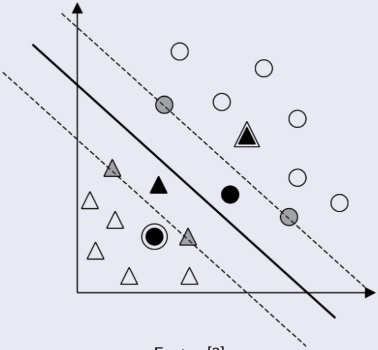


Fonte: [3]

SVM Linear com Margens Suaves

Dados com ruído

- O classificador permite que alguns exemplos caiam no lado errado do limite de decisão
  - Associa, contudo, uma penalidade proporcional à distância necessária para movê-los de volta ao lado certo

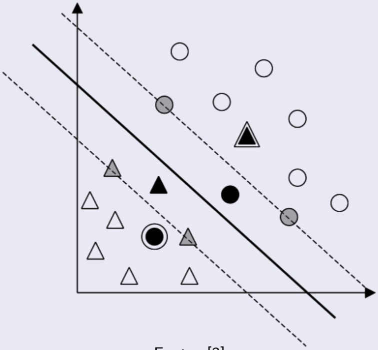


Fonte: [3]

SVM Linear com Margens Suaves

Dados com ruído

- O resultado é a mesma expressão para o classificador com margens rígidas
  - Mas com uma expressão diferente para α_i (e consequentes ω e b)
  - Não veremos detalhes aqui

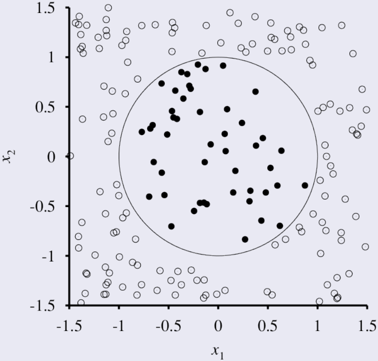


Fonte: [3]

SVM Não Linear

Espaço de Características

- E se os dados não forem linearmente separáveis?
  - Muito embora, nesse exemplo, talvez bastasse o uso de coordenadas polares

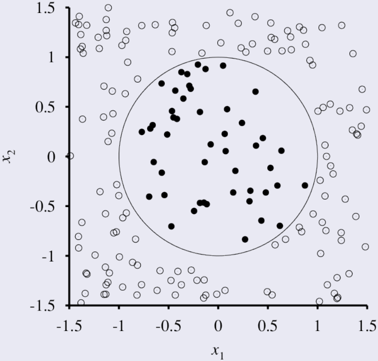


Fonte: AIMA. R&N

SVM Não Linear

Espaço de Características

- E se os dados não forem linearmente separáveis?
  - Muito embora, nesse exemplo, talvez bastasse o uso de coordenadas polares
- SVMs lidam com isso mapeando cada exemplo para um novo espaço, de maior dimensão
  - O **espaço de características** (*feature space*)

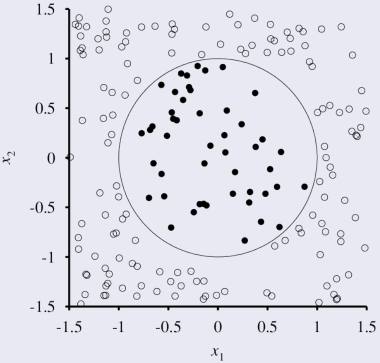


Fonte: AIMA. R&N

SVM Não Linear

Espaço de Características

- Por exemplo, vamos mapear cada vetor de entrada $\vec{x} = (x_1, x_2)$ em um novo vetor $F(\vec{x}) = (f_1, f_2, f_3)$, onde:
  - $f_1 = x_1^2$
  - $f_2 = x_2^2$
  - $f_3 = \sqrt{2}x_1x_2$

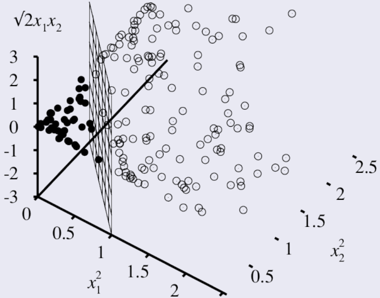


Fonte: AIMA. R&N

SVM Não Linear

Espaço de Características

- Graficando os dados nesse novo espaço obtemos

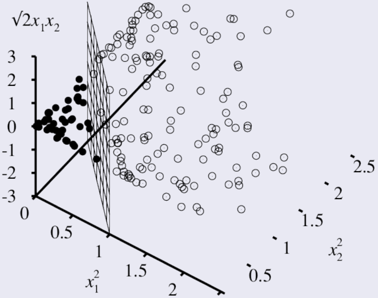


Fonte: AIMA. R&N

SVM Não Linear

Espaço de Características

- Graficando os dados nesse novo espaço obtemos
- E eles são linearmente separáveis
  - A escolha apropriada do mapeamento faz com que os dados possam ser separados por uma SVM linear

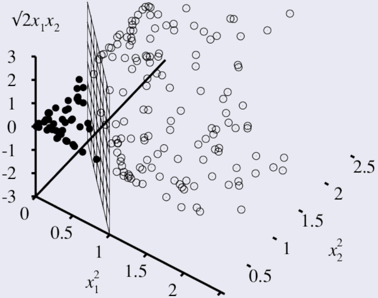


Fonte: AIMA. R&N

SVM Não Linear

Espaço de Características

- Se os dados forem mapeados em um espaço de dimensões suficientemente grande, eles quase sempre serão linearmente separáveis
  - Se olharmos a um conjunto de pontos a partir de direções suficientes, encontraremos um modo de alinhá-los



Fonte: AIMA. R&N

  - Com algumas exceções, conjuntos de n pontos serão sempre separáveis em espaços de $n - 1$ dimensões ou mais

SVM Não Linear

Separador linear

- Para encontrar um separador linear no novo espaço $F(\vec{x})$, substituímos $\vec{x}_i \cdot \vec{x}_j$ por $F(\vec{x}_i) \cdot F(\vec{x}_j)$ em

$$\max_{\vec{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$$

- Obtendo

$$\max_{\vec{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (F(\vec{x}_i) \cdot F(\vec{x}_j))$$

SVM Não Linear

Funções de Kernel

- Contudo, podemos calcular $F(\vec{x}_i) \cdot F(\vec{x}_j)$ sem ter de calcular F para cada ponto
  - No exemplo dado,

$$\begin{aligned} F(\vec{x}_i) \cdot F(\vec{x}_j) &= (x_{1i}^2, \sqrt{2}x_{1i}x_{2i}, x_{2i}^2) \cdot (x_{1j}^2, \sqrt{2}x_{1j}x_{2j}, x_{2j}^2) \\ &= (\vec{x}_i \cdot \vec{x}_j)^2 \end{aligned}$$

SVM Não Linear

Funções de Kernel

- Contudo, podemos calcular $F(\vec{x}_i) \cdot F(\vec{x}_j)$ sem ter de calcular F para cada ponto
  - No exemplo dado,

$$\begin{aligned} F(\vec{x}_i) \cdot F(\vec{x}_j) &= (x_{1i}^2, \sqrt{2}x_{1i}x_{2i}, x_{2i}^2) \cdot (x_{1j}^2, \sqrt{2}x_{1j}x_{2j}, x_{2j}^2) \\ &= (\vec{x}_i \cdot \vec{x}_j)^2 \end{aligned}$$

- $K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j)^2$ é uma **Função de Kernel**
  - Uma função $K(\vec{x}_i, \vec{x}_j) = F(\vec{x}_i) \cdot F(\vec{x}_j)$ que recebe 2 pontos $\vec{x}_i$ e $\vec{x}_j$ do espaço de entradas e calcula seu produto escalar no espaço de características

SVM Não Linear

Funções de Kernel

- Podemos então encontrar separadores lineares em $F(\vec{x})$ simplesmente trocando $\vec{x}_i \cdot \vec{x}_j$ pela função de kernel $K(\vec{x}_i, \vec{x}_j)$
  - Para aprender em dimensões maiores, calculamos apenas as funções de kernel, em vez da lista de características inteira para cada ponto

SVM Não Linear

Funções de Kernel

- Podemos então encontrar separadores lineares em $F(\vec{x})$ simplesmente trocando $\vec{x}_i \cdot \vec{x}_j$ pela função de kernel $K(\vec{x}_i, \vec{x}_j)$
  - Para aprender em dimensões maiores, calculamos apenas as funções de kernel, em vez da lista de características inteira para cada ponto
- Assim, simplificamos o cálculo
  - Empregamos a função de Kernel sem conhecer o mapeamento F, pois esse é usado implicitamente
  - Podemos encontrar separadores lineares eficientemente em espaços de bilhões de dimensões

SVM Não Linear

Funções de Kernel

- E podemos usar qualquer função como Kernel?

SVM Não Linear

Funções de Kernel

- E podemos usar qualquer função como Kernel?
  - Não. Apenas funções que satisfaçam as condições estabelecidas pelo teorema de Mercer

SVM Não Linear

Funções de Kernel

- E podemos usar qualquer função como Kernel?
  - Não. Apenas funções que satisfaçam as condições estabelecidas pelo teorema de Mercer
- Um Kernel que satisfaz as condições de Mercer dá origem a matrizes positivas semi-definidas $[K]$
  - Em que cada elemento K_{ij} é definido como $K_{ij} = K(\vec{x}_i, \vec{x}_j)$, para $i, j = 1, \ldots, n$

SVM Não Linear

Funções de Kernel

- Na prática, os Kernels mais usados são

Tipo	$K(\vec{x}_i, \vec{x}_j)$	*Parâmetros*
Linear	$\delta(\vec{x}_i \cdot \vec{x}_j) + \kappa$	δ e κ
Polinomial	$(\delta(\vec{x}_i \cdot \vec{x}_j) + \kappa)^d$	δ, κ e d
Gaussiano	$e^{-\sigma\|\vec{x}_i - \vec{x}_j\|^2}$	σ
Sigmoidal	$tanh(\delta(\vec{x}_i \cdot \vec{x}_j) + \kappa)$	δ e κ

- Note que cada um deles apresenta hiper-parâmetros que precisam ser determinados na prática
- No caso do sigmoidal, as condições de Mercer são satisfeitas somente para alguns valores de δ

Support Vector Machines

Vantagens

- Constroem um separador de margem máxima
  - Um limite de decisão com a maior distância possível dos exemplos de treino, o que ajuda a generalizar o modelo

Support Vector Machines

Vantagens

- Constroem um separador de margem máxima
  - Um limite de decisão com a maior distância possível dos exemplos de treino, o que ajuda a generalizar o modelo
- Criam um plano de separação linear
  - Tornam isso possível embutindo os dados em um espaços de mais dimensões (via o uso de *Kernels*)
  - Frequentemente, dados não linearmente separáveis no espaço original se tornam separáveis nesse espaço maior
  - O separador linear de alta dimensão não é linear no espaço original $\rightarrow$ podemos representar hipóteses não lineares

Support Vector Machines

Vantagens

- São não-paramétricos
  - Retêm exemplos e potencialmente precisam armazená-los todos
  - Na prática, contudo, apenas retêm uma fração pequena destes

Support Vector Machines

Vantagens

- São não-paramétricos
  - Retêm exemplos e potencialmente precisam armazená-los todos
  - Na prática, contudo, apenas retêm uma fração pequena destes
- Combinam assim as vantagens dos modelos não-paramétricos e paramétricos
  - Possuem a flexibilidade para representar funções complexas
  - E ainda assim são resistentes a *overfitting*

Support Vector Machines

Desvantagens

- Sensíveis à escolha dos parâmetros
- Sensíveis à escolha do *Kernel*
- Nativamente só tratam de classificação binária

Referências

1. Russell, S.; Norvig P. (2010): Artificial Intelligence: A Modern Approach. Prentice Hall. 2a e 3a ed.
2. Harrington, P. (2012): Machine Learning in Action. Manning.
3. Lorena, A.C.; Carvalho A.C.P.L.F. (2007): Uma Introdução às Support Vector Machines. RITA, 14(2).
4. Haykin, S. (2009): Neural Networks and Learning Machines. Pearson. 3 ed.
5. `https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/`