

Classificação

ACH5504 – Mineração de Dados

Notas de aulas baseadas no livro

“Introduction to Data Mining”

Tan, Steinbach, Karpatne, Kumar

Resumo

- Definição de classificação
- Naive Bayes
- Support Vector Machines
- Assembleias
- Regras

Definição de classificação

- Para uma coleção de registros (conjunto de treinamento)
 - Cada registro é caracterizado por uma tupla (x,y) , onde x é o conjunto de atributos e y é o rótulo da classe
 - x : atributo, preditor, variável independente, entrada
 - y : classe, resposta, variável dependente, saída
- Tarefa:
 - Aprender um modelo que mapeia cada atributo definido como x em um dos rótulos de classe predefinidos y

Classificador de Bayes

- Uma estrutura probabilística para resolver problemas de classificação

- Probabilidade condicional:
$$P(Y | X) = \frac{P(X, Y)}{P(X)}$$

$$P(X | Y) = \frac{P(X, Y)}{P(Y)}$$

- Teorema de Bayes:

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

Exemplo de teorema de Bayes

- Dados:
 - Um médico sabe que a meningite provoca rigidez no pescoço 50% do tempo, $P(S/M) = 0.5$
 - A probabilidade prévia de qualquer paciente ter meningite é $P(M) = 1/50000$
 - A probabilidade prévia de qualquer paciente com torcicolo é $P(S) = 1/20$
- Se um paciente tem torcicolo, qual é a probabilidade de ter meningite?

$$P(M | S) = \frac{P(S | M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

Usando o teorema de Bayes para classificação

- Considere cada atributo e rótulo de classe como variáveis aleatórias
- Dado um registro com atributos (X_1, X_2, \dots, X_d)
 - O objetivo é prever a classe Y
 - Especificamente, queremos encontrar o valor de Y que maximiza a probabilidade $P(Y | X_1, X_2, \dots, X_d)$
- Podemos estimar $P(Y | X_1, X_2, \dots, X_d)$ diretamente dos dados?

Dados exemplares

Dado um registro de teste:

$$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Podemos estimar $P(\text{Evade} = \text{Yes} \mid X)$ e $P(\text{Evade} = \text{No} \mid X)$?

A seguir, substituiremos

Evade = Yes por Yes, e

Evade = No por No

Usando o teorema de Bayes para classificação

- Abordagem:
 - Calcular a probabilidade posterior $P(Y | X_1, X_2, \dots, X_d)$ usando o teorema de Bayes

$$P(Y | X_1 X_2 \dots X_n) = \frac{P(X_1 X_2 \dots X_d | Y) P(Y)}{P(X_1 X_2 \dots X_d)}$$

- *O máximo a-posteriori*: escolha Y que maximiza $P(Y | X_1, X_2, \dots, X_d)$
 - Equivalente a escolher o valor de Y que maximiza $P(X_1, X_2, \dots, X_d | Y) P(Y)$
- Como estimar $P(X_1, X_2, \dots, X_d | Y)$?

Dados exemplares

Dado um registro de teste:

$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Usando o teorema de Bayes:

- $$P(\text{Yes}|X) = \frac{P(X|\text{Yes})P(\text{Yes})}{P(X)}$$

- $$P(\text{No}|X) = \frac{P(X|\text{No})P(\text{No})}{P(X)}$$

- Como estimar $P(X|\text{Yes})$ e $P(X|\text{No})$?

Classificador de Naïve Bayes

- Suponha a independência entre os atributos X_i quando a classe é dada:
 - $P(X_1, X_2, \dots, X_d | Y_j) = P(X_1 | Y_j) P(X_2 | Y_j) \dots P(X_d | Y_j)$
 - Agora podemos estimar $P(X_i | Y_j)$ para todas combinações de X_i e Y_j usando dados de treino
 - Novo ponto é classificado para Y_j se $P(Y_j) \prod P(X_i | Y_j)$ é o máximo.

Independência condicional

- **X** e **Y** são condicionalmente independentes dado **Z** se $P(\mathbf{X}|\mathbf{YZ}) = P(\mathbf{X}|\mathbf{Z})$
- Exemplo: Comprimento do braço e habilidades de leitura
 - A criança jovem tem um comprimento mais curto do braço e umas habilidades limitadas da leitura, comparadas aos adultos
 - Se a idade é fixa, nenhuma relação aparente entre o comprimento do braço e as habilidades de leitura
 - Comprimento do braço e habilidades de leitura são condicionalmente independente dada idade

Naïve Bayes com dados exemplares

Dado um registo de teste:

$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- $P(X | \text{Yes}) =$
 $P(\text{Refund} = \text{No} | \text{Yes}) \times$
 $P(\text{Divorced} | \text{Yes}) \times$
 $P(\text{Income} = 120\text{K} | \text{Yes})$
- $P(X | \text{No}) =$
 $P(\text{Refund} = \text{No} | \text{No}) \times$
 $P(\text{Divorced} | \text{No}) \times$
 $P(\text{Income} = 120\text{K} | \text{No})$

Estimativa de probabilidades usando dados

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Classe: $P(Y) = N_c/N$
 - e.g., $P(\text{No}) = 7/10$,
 $P(\text{Yes}) = 3/10$
- Para atributos categóricos:
$$P(X_i | Y_k) = |X_{ik}| / N_{c,k}$$
 - onde $|X_{ik}|$ é o número de instâncias com valor de atributo X_i fazendo parte da classe Y_k
 - Exemplos:
 $P(\text{Status}=\text{Married} | \text{No}) = 4/7$
 $P(\text{Refund}=\text{Yes} | \text{Yes})=0$

Estimativa de probabilidades usando dados

- Para atributos contínuos:
 - **Discretização:** Particionar o intervalo em compartimentos:
 - ◆ Substitua o valor contínuo pelo valor do compartimento
 - Atributo alterado de contínuo para ordinal
 - **Estimativa da densidade de probabilidade:**
 - ◆ Assume que o atributo segue uma distribuição normal
 - ◆ Use dados para estimar parâmetros de distribuição (e.g., média e desvio padrão)
 - ◆ Uma vez que a distribuição de probabilidade é conhecida, use-a para estimar a probabilidade condicional $P(X_i|Y)$

Estimativa de probabilidades usando dados

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Distribuição normal:

$$P(X_i | Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(X_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

- Um para cada par (X_i, Y_i)

- Para (Income, Classe = No):

- Se Classe = No

- ◆ média da amostra = 110
- ◆ variância da amostra = 2975

$$P(\text{Income} = 120 | \text{No}) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Exemplo de Classificador Naïve Bayes

Dado um registro de teste:

$$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$$

Classificador Naïve Bayes:

$$P(\text{Refund} = \text{Yes} \mid \text{No}) = 3/7$$

$$P(\text{Refund} = \text{No} \mid \text{No}) = 4/7$$

$$P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0$$

$$P(\text{Refund} = \text{No} \mid \text{Yes}) = 1$$

$$P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/7$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 1/7$$

$$P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/7$$

$$P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3$$

$$P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0$$

Para "Taxable Income":

Se classe = No: média = 110

variância = 2975

Se classe = Yes: média = 90

variância = 25

- $P(X \mid \text{No}) = P(\text{Refund}=\text{No} \mid \text{No})$
 $\times P(\text{Divorced} \mid \text{No})$
 $\times P(\text{Income}=120\text{K} \mid \text{No})$
 $= 4/7 \times 1/7 \times 0.0072 = 0.0006$
- $P(X \mid \text{Yes}) = P(\text{Refund}=\text{No} \mid \text{Yes})$
 $\times P(\text{Divorced} \mid \text{Yes})$
 $\times P(\text{Income}=120\text{K} \mid \text{Yes})$
 $= 1 \times 1/3 \times 1.2 \times 10^{-9} = 4 \times 10^{-10}$

Desde $P(X|\text{No})P(\text{No}) > P(X|\text{Yes})P(\text{Yes})$

então $P(\text{No}|X) > P(\text{Yes}|X)$

=> Classe = No

Exemplo de Classificador Naïve Bayes

Dado um registro de teste:

$$X = (\text{Refund} = \text{No}, \text{Divorced}, \text{Income} = 120\text{K})$$

Classificador Naïve Bayes:

$$P(\text{Refund} = \text{Yes} \mid \text{No}) = 3/7$$

$$P(\text{Refund} = \text{No} \mid \text{No}) = 4/7$$

$$P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0$$

$$P(\text{Refund} = \text{No} \mid \text{Yes}) = 1$$

$$P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/7$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 1/7$$

$$P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/7$$

$$P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3$$

$$P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0$$

Para "Taxable Income":

Se classe = No: média = 110

variância = 2975

Se classe = Yes: média = 90

variância = 25

- $P(\text{Yes}) = 3/10$

$$P(\text{No}) = 7/10$$

- $P(\text{Yes} \mid \text{Divorced}) = (1/3 \times 3/10) / P(\text{Divorced})$

$$P(\text{No} \mid \text{Divorced}) = (1/7 \times 7/10) / P(\text{Divorced})$$

- $P(\text{Yes} \mid \text{Refund} = \text{No}, \text{Divorced}) = 1 \times 1/3 \times 3/10 / P(\text{Divorced}, \text{Refund} = \text{No})$

$$P(\text{No} \mid \text{Refund} = \text{No}, \text{Divorced}) = 4/7 \times 1/7 \times 7/10 / P(\text{Divorced}, \text{Refund} = \text{No})$$

Problemas com o classificador Naïve Bayes

Classificador Naïve Bayes:

$$P(\text{Refund} = \text{Yes} \mid \text{No}) = 3/7$$

$$P(\text{Refund} = \text{No} \mid \text{No}) = 4/7$$

$$P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0$$

$$P(\text{Refund} = \text{No} \mid \text{Yes}) = 1$$

$$P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/7$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 1/7$$

$$P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/7$$

$$P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3$$

$$P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0$$

Para “Taxable Income”:

Se classe = No: média = 110

variância = 2975

Se classe = Yes: média = 90

variância = 25

- $P(\text{Yes}) = 3/10$

$$P(\text{No}) = 7/10$$

- $P(\text{Yes} \mid \text{Married}) = (0 \times 3/10) / P(\text{Married})$

$$P(\text{No} \mid \text{Married}) = (4/7 \times 7/10) / P(\text{Married})$$

Problemas com o classificador Naïve Bayes

Considere a tabela com Tid = 7 excluído

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Classificador Naïve Bayes:

$$P(\text{Refund} = \text{Yes} \mid \text{No}) = 2/6$$

$$P(\text{Refund} = \text{No} \mid \text{No}) = 4/6$$

$$\rightarrow P(\text{Refund} = \text{Yes} \mid \text{Yes}) = 0$$

$$P(\text{Refund} = \text{No} \mid \text{Yes}) = 1$$

$$P(\text{Marital Status} = \text{Single} \mid \text{No}) = 2/6$$

$$\rightarrow P(\text{Marital Status} = \text{Divorced} \mid \text{No}) = 0$$

$$P(\text{Marital Status} = \text{Married} \mid \text{No}) = 4/6$$

$$P(\text{Marital Status} = \text{Single} \mid \text{Yes}) = 2/3$$

$$P(\text{Marital Status} = \text{Divorced} \mid \text{Yes}) = 1/3$$

$$P(\text{Marital Status} = \text{Married} \mid \text{Yes}) = 0/3$$

Para "Taxable Income":

Se classe = No: média = 91
variância = 685

If class = No: média = 90
variância = 25

Dado X = (Refund = Yes, Divorced, 120K)

$$P(X \mid \text{No}) = 2/6 \times 0 \times 0.0083 = 0$$

$$P(X \mid \text{Yes}) = 0 \times 1/3 \times 1.2 \times 10^{-9} = 0$$

Naïve Bayes não será capaz de classificar X como Yes ou No!

Problemas com o classificador Naïve Bayes

- Se uma das probabilidades condicionais for zero, a expressão inteira se tornará zero
- Necessidade de usar outras estimativas de probabilidades condicionais do que frações simples
- Estimativa de probabilidade:

$$\text{Original: } P(A_i | C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace: } P(A_i | C) = \frac{N_{ic} + 1}{N_c + c}$$

$$\text{m - estimate: } P(A_i | C) = \frac{N_{ic} + mp}{N_c + m}$$

c : número de classes

p : probabilidade prévia da classe

m : parâmetro

N_c : número de instâncias na classe

N_{ic} : número de instâncias com valor de atributo A_i na classe c

Exemplo de classificador Naïve Bayes

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: atributos

M: mamíferos

N: não mamíferos

$$P(A | M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A | N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A | M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A | N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

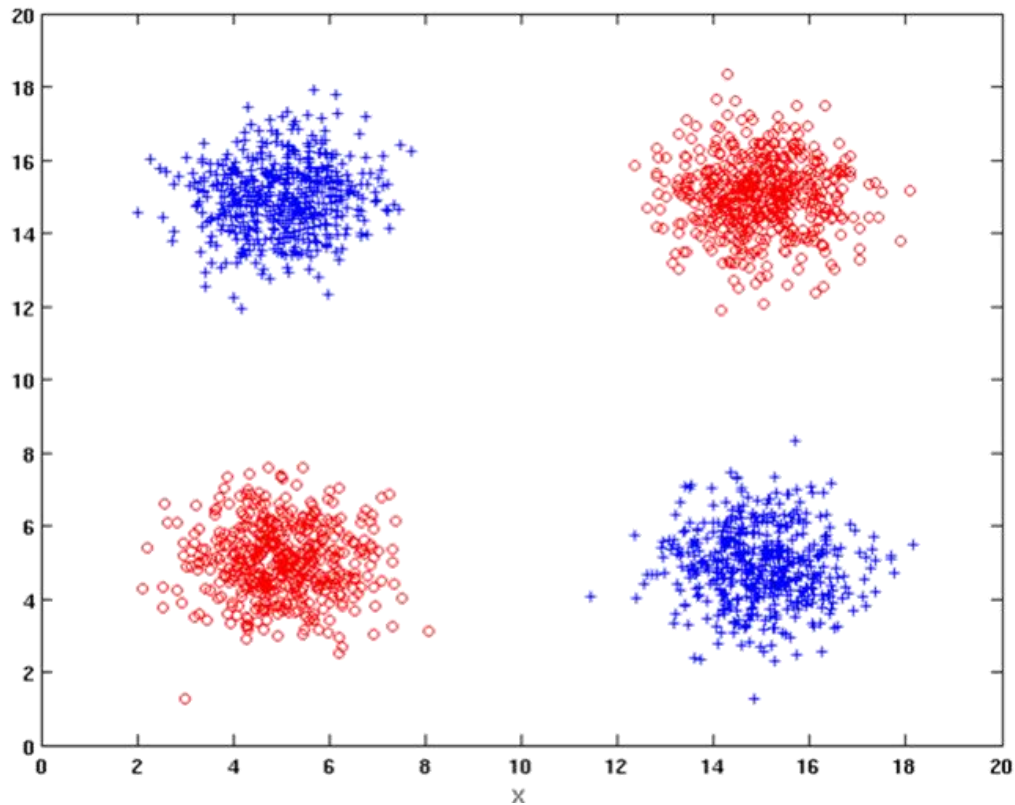
=> Mamíferos

Sumário de Naïve Bayes

- Robusto para pontos de ruído isolados
- Trabalha com valores ausentes ignorando a instância durante calculo de estimativa de probabilidade
- Robusto a atributos irrelevantes
- A suposição de independência pode não ser válida para alguns atributos
 - Use outra técnica como Bayesian Belief Networks (BBN)

Naïve Bayes

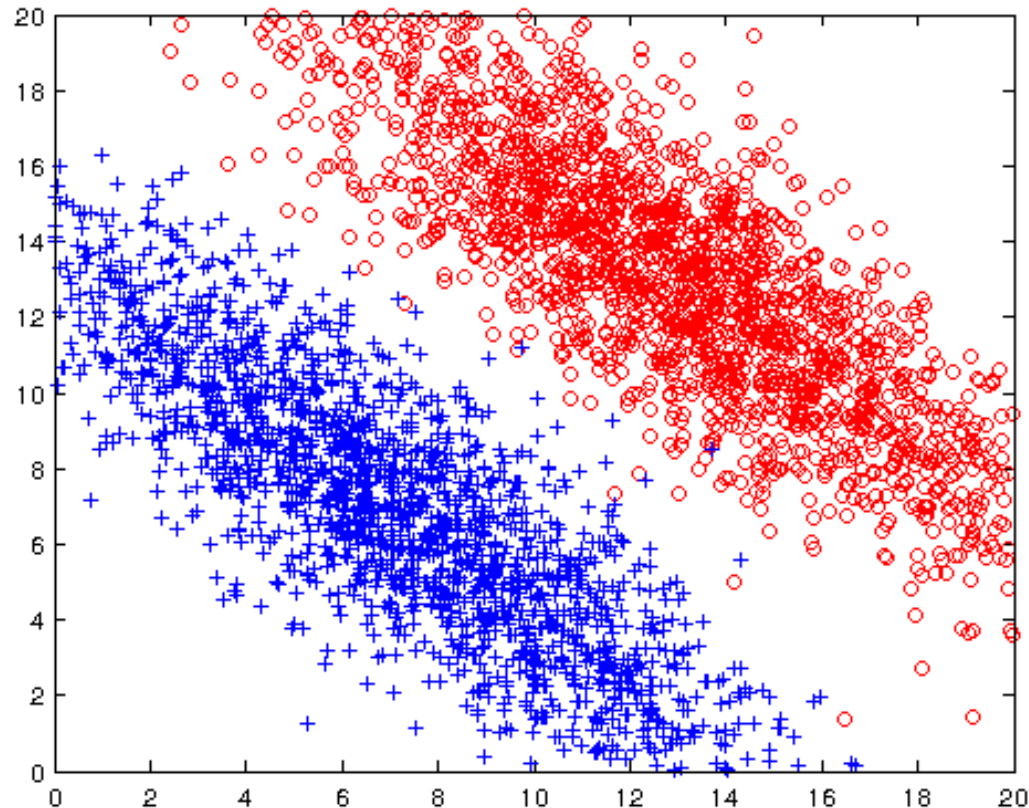
- Como o Naïve Bayes funciona no seguinte conjunto de dados?



A independência condicional dos atributos é violada

Naïve Bayes

- Como o Naïve Bayes funciona no seguinte conjunto de dados?



Naïve Bayes pode construir limites de decisão oblíqua

Naïve Bayes

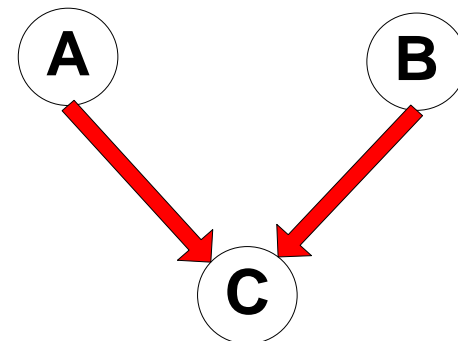
- Como o Naïve Bayes funciona no seguinte conjunto de dados?

Y = 1	1	1	1	0
Y = 2	0	1	0	0
Y = 3	0	0	1	1
Y = 4	0	0	1	1
	X = 1	X = 2	X = 3	X = 4

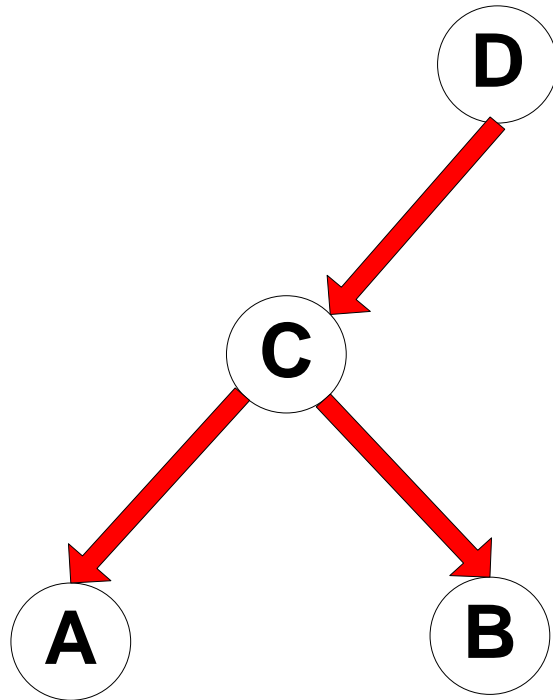
A independência condicional dos atributos é violada

Bayesian Belief Networks

- Fornece representação gráfica de relações probabilísticas entre um conjunto de variáveis aleatórias
- Consiste em:
 - Um gráfico acíclico direcionado (directed acyclic graph - dag)
 - ◆ Nó corresponde a uma variável
 - ◆ Seta corresponde à relação de dependência entre um par de variáveis
 - Uma tabela de probabilidade que associe cada nó ao seu pai imediato



Independência condicional



D é pai de C

A é filho de C

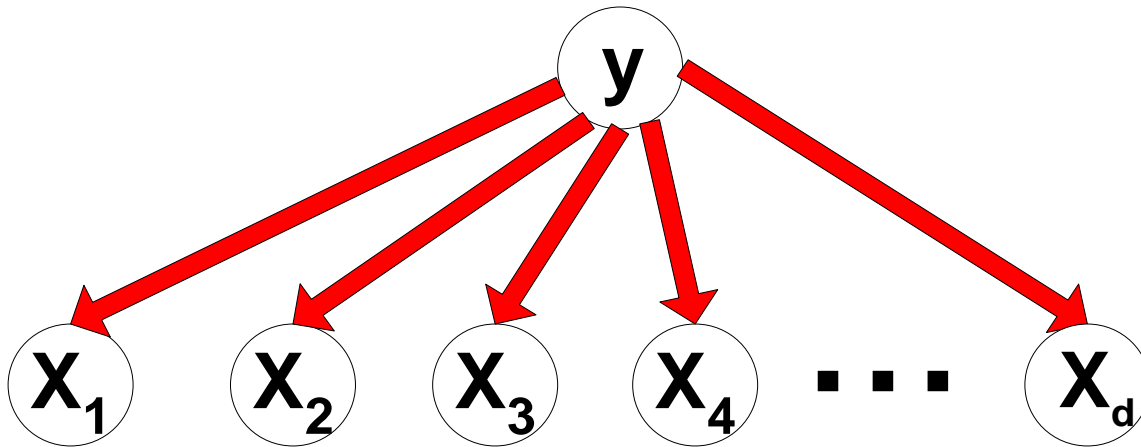
B é descendente de D

D é ancestral de A

- Um nó em uma rede bayesiana é condicionalmente independente de todos os seus não descendentes, se seus pais são conhecidos

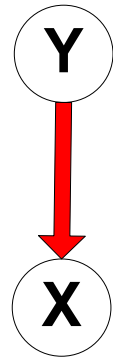
Independência condicional

- Suposição de Naïve Bayes:



Tabelas de probabilidade

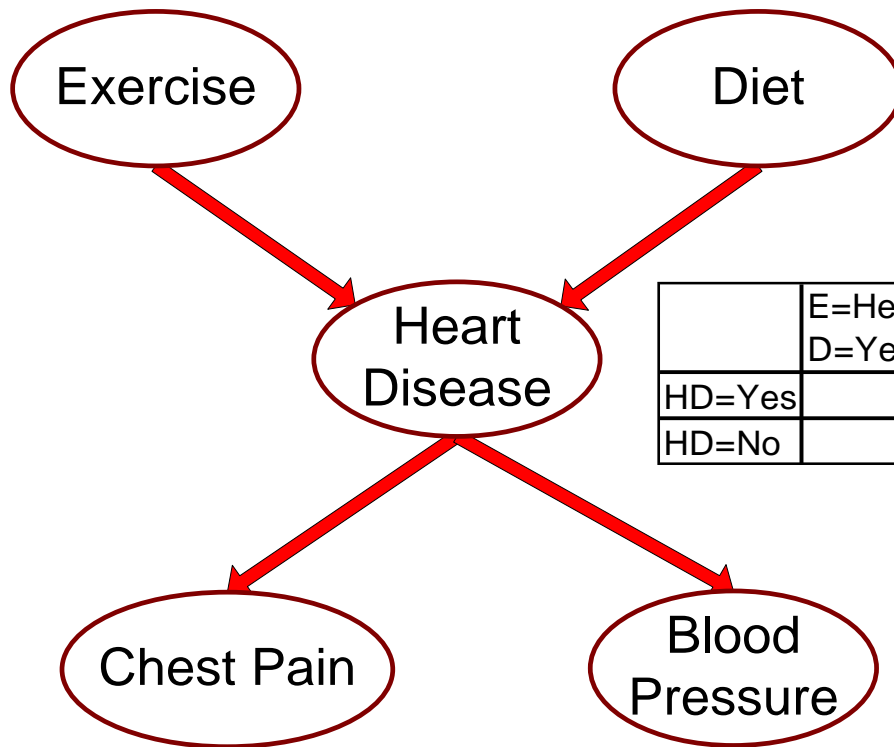
- Se X não tiver pais, a tabela contém probabilidade prévia $P(X)$.
- Se X tiver apenas um pai (Y), a tabela conterá a probabilidade condicional $P(X|Y)$.
- Se X tiver vários pais (Y_1, Y_2, \dots, Y_k), a tabela conterá a probabilidade condicional $P(X|Y_1, Y_2, \dots, Y_k)$



Exemplo de Bayesian Belief Network

Exercise=Yes	0.7
Exercise=No	0.3

Diet=Healthy	0.25
Diet=Unhealthy	0.75

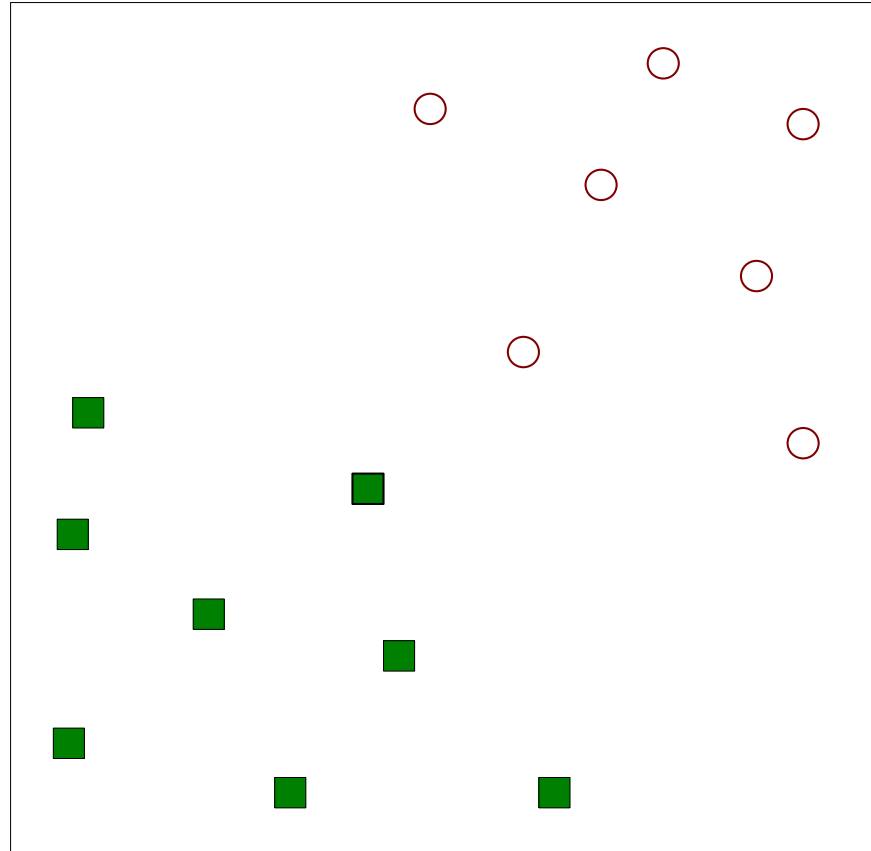


	E=Healthy D=Yes	E=Healthy D=No	E=Unhealthy D=Yes	E=Unhealthy D=No
HD=Yes	0.25	0.45	0.55	0.75
HD=No	0.75	0.55	0.45	0.25

	HD=Yes	HD=No
CP=Yes	0.8	0.01
CP=No	0.2	0.99

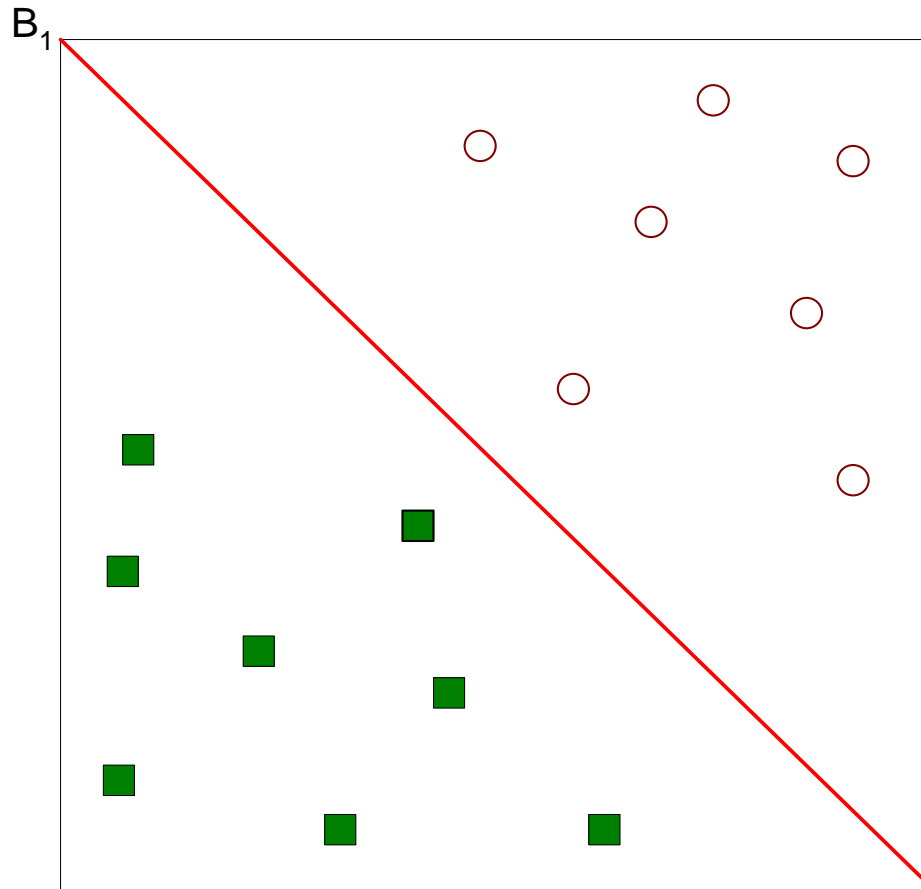
	HD=Yes	HD=No
BP=High	0.85	0.2
BP=Low	0.15	0.8

Support Vector Machines



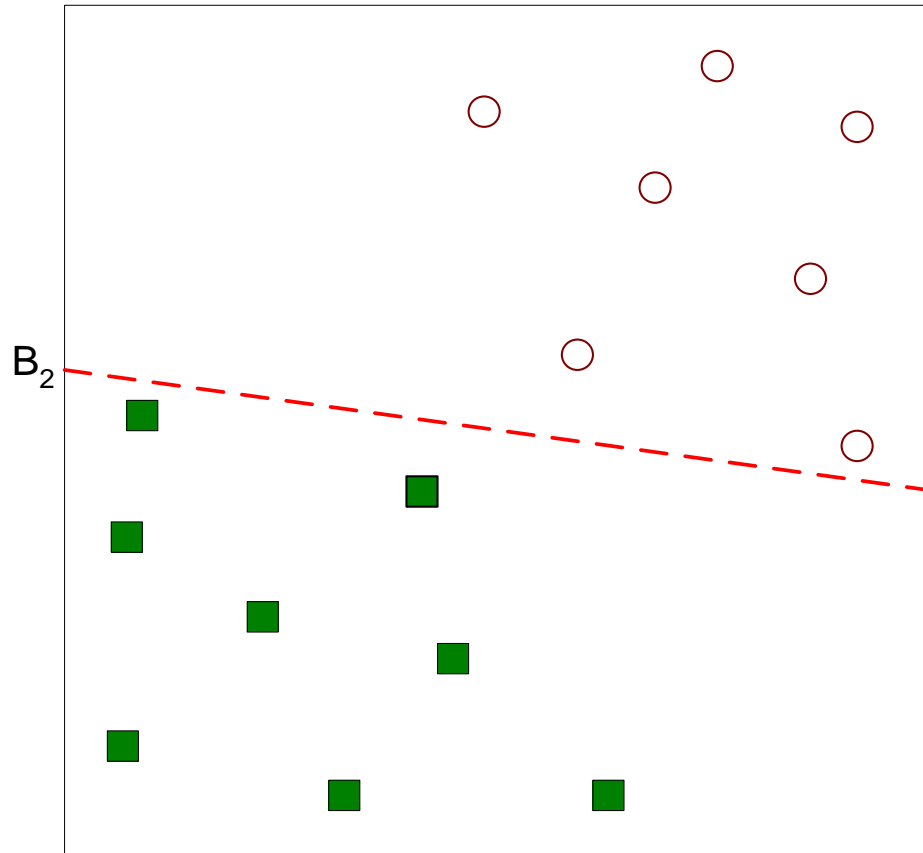
- Encontre um hiperplano linear (limite de decisão) que separará os dados

Support Vector Machines



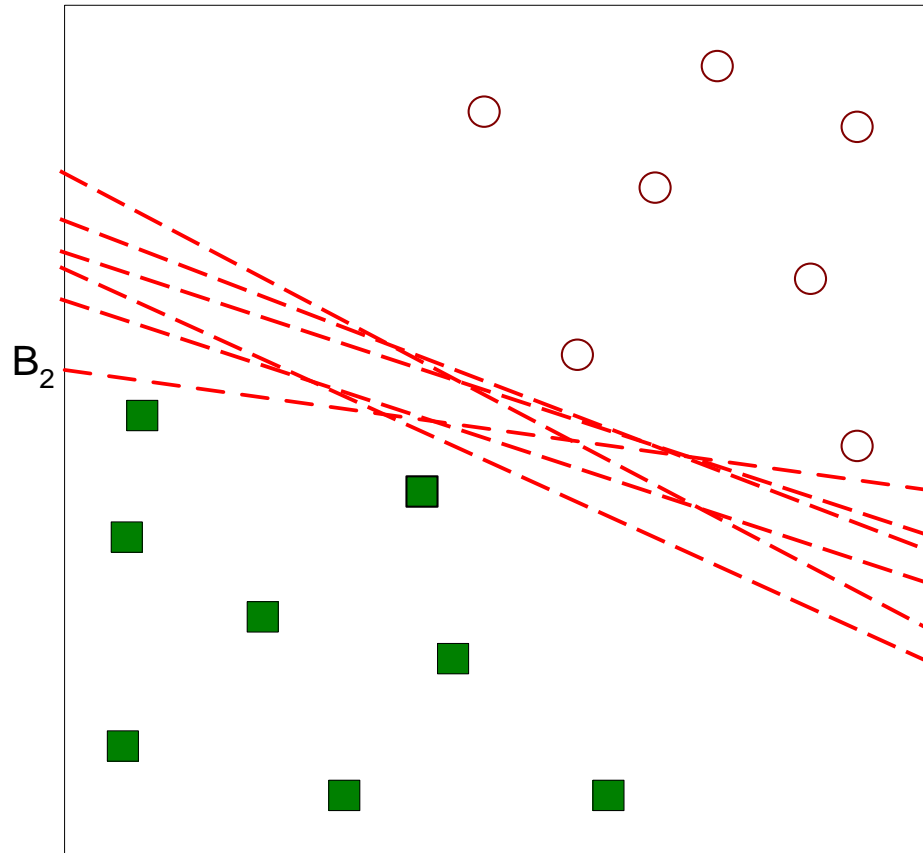
- Uma solução possível

Support Vector Machines



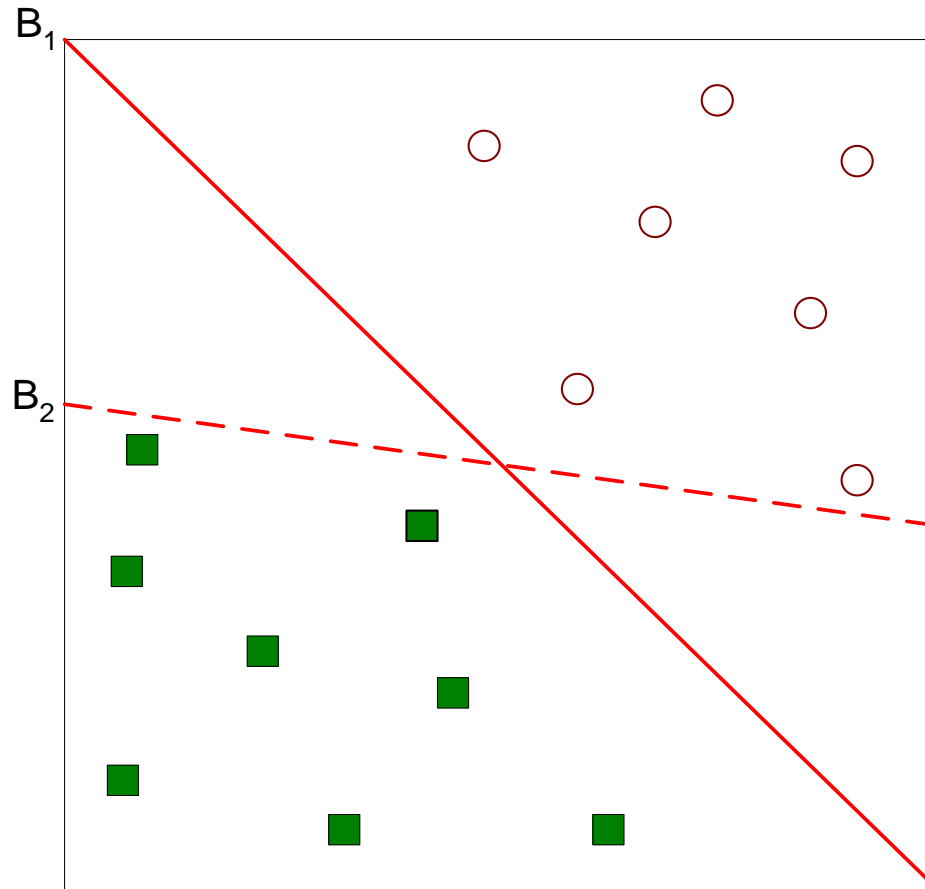
- Outra solução possível

Support Vector Machines



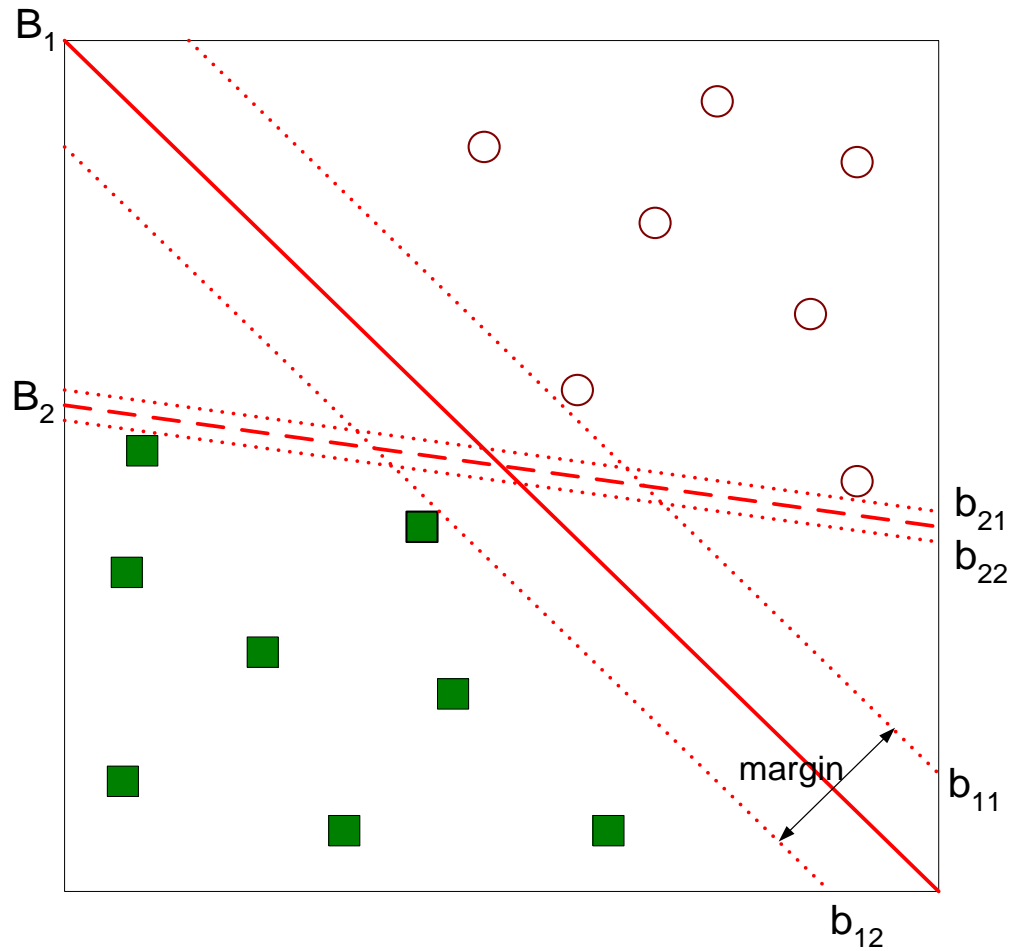
- Outras soluções possíveis

Support Vector Machines



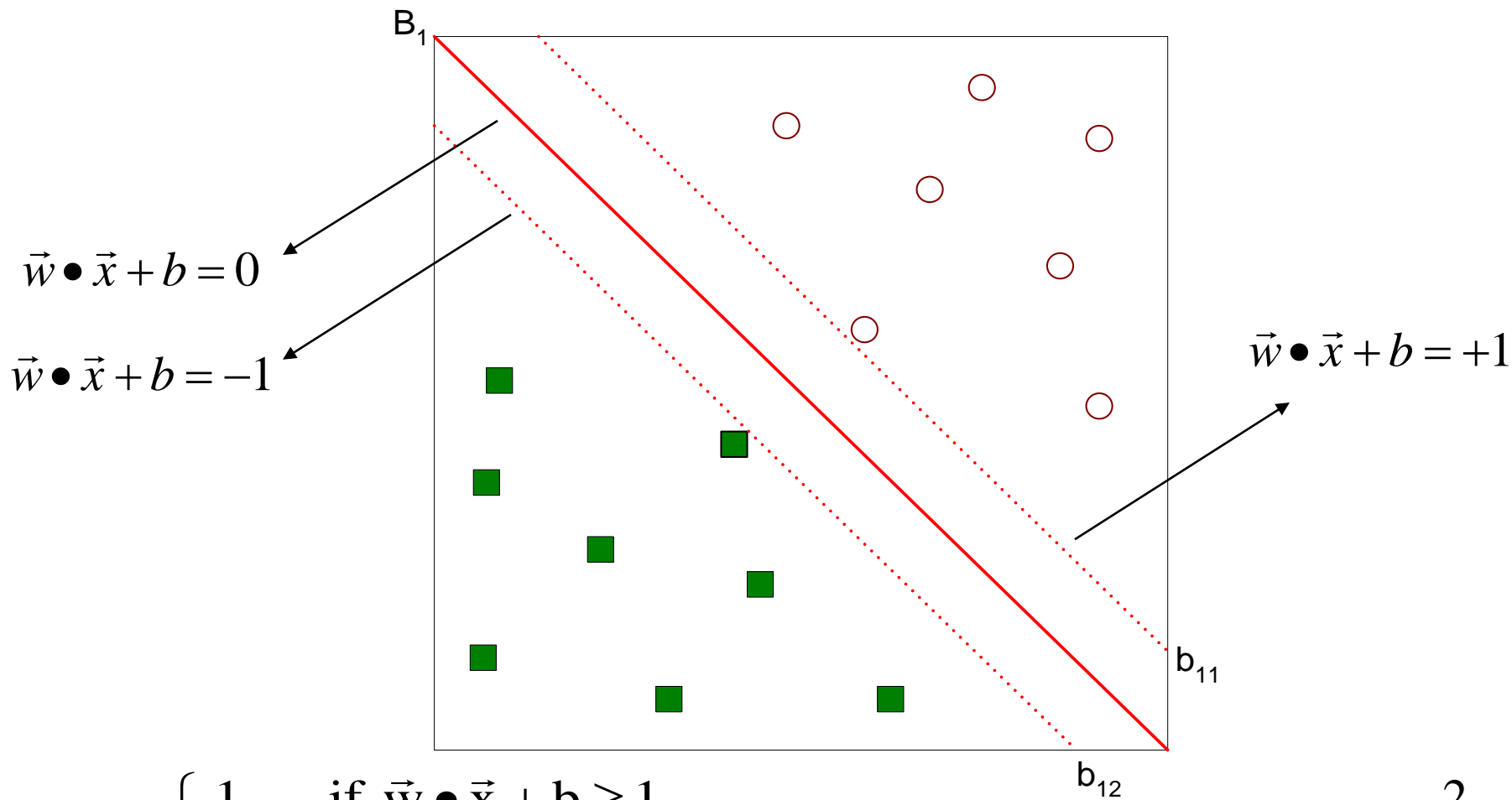
- Qual delas é melhor? B_1 ou B_2 ?
- Como você define a melhor divisão?

Support Vector Machines



- Encontre o hiperplano que **maximiza** a margem => B_1 é melhor do que B_2

Support Vector Machines



$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

$$\text{Margin} = \frac{2}{\|\vec{w}\|}$$

SVM Linear

- Modelo linear:

$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

- Ensinar o modelo equivale a determinar os valores de **w** e **b**
 - Como achar estes valores de dados de treino?

Aprendizagem de SVM Linear

- O objetivo é maximizar: $\text{Margin} = \frac{2}{\|\vec{w}\|}$

– O que equivale a minimizar:

$$L(\vec{w}) = \frac{\|\vec{w}\|^2}{2}$$

– Sujeito às seguintes restrições:

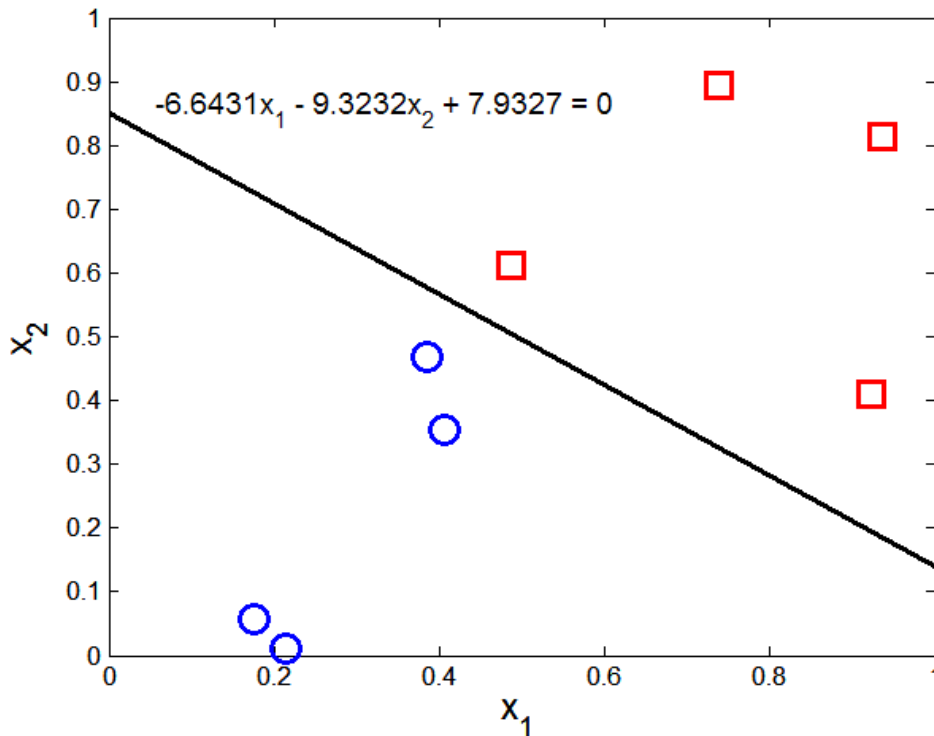
ou
$$y_i = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x}_i + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x}_i + b \leq -1 \end{cases}$$

$$y_i(\mathbf{w} \bullet \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, N$$

- Este é um problema de otimização restrita
 - Resolvê-lo usando o método multiplicador de Lagrange

$$L_p = \frac{\|\vec{w}\|^2}{2} - \sum_{i=1}^N \lambda_i (y_i(\vec{w} \cdot \vec{x}_i + b) - 1)$$

Exemplo de SVM Linear



Lagrange multipliers
ou Support vectors

x1	x2	y	λ
0.3858	0.4687	1	65.5261
0.4871	0.611	-1	65.5261
0.9218	0.4103	-1	0
0.7382	0.8936	-1	0
0.1763	0.0579	1	0
0.4057	0.3529	1	0
0.9355	0.8132	-1	0
0.2146	0.0099	1	0

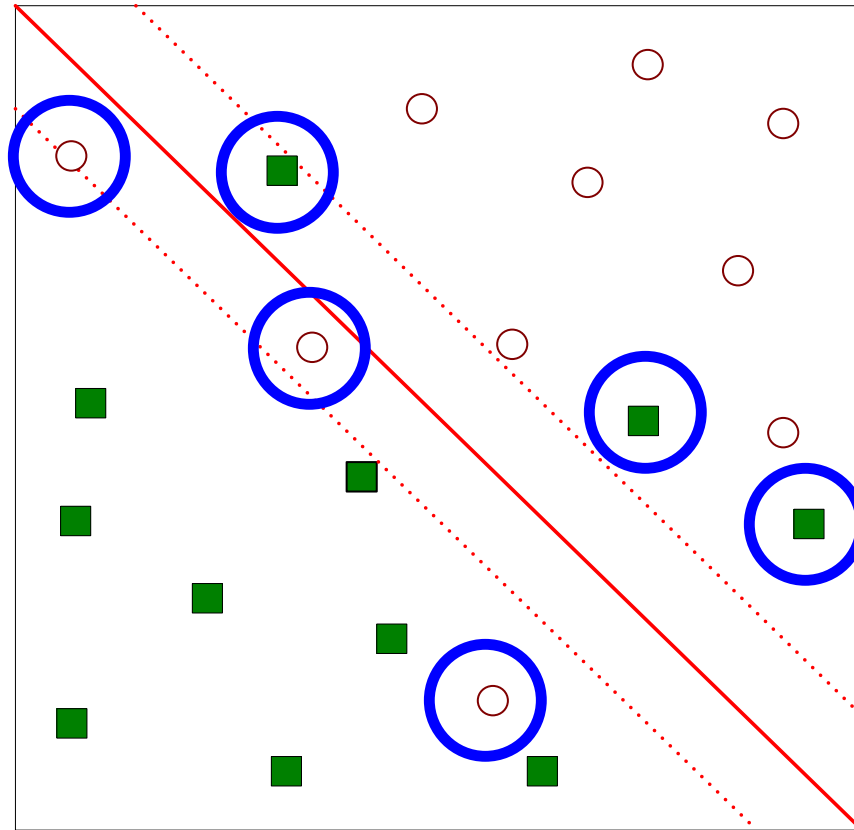
Aprendizagem de SVM Linear

- Limite de decisão depende apenas de vetores de suporte
 - Se você tiver conjunto de dados com os mesmos vetores de suporte, o limite de decisão não será alterado.
 - Como classificar usando SVM uma vez \mathbf{w} e b são encontrados? Dado um registro de teste, x_i

$$f(\vec{x}_i) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x}_i + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x}_i + b \leq -1 \end{cases}$$

Support Vector Machines

- E se o problema não for linearmente separável?



Support Vector Machines

- E se o problema não for linearmente separável?
 - Introduzir variáveis de “folga”

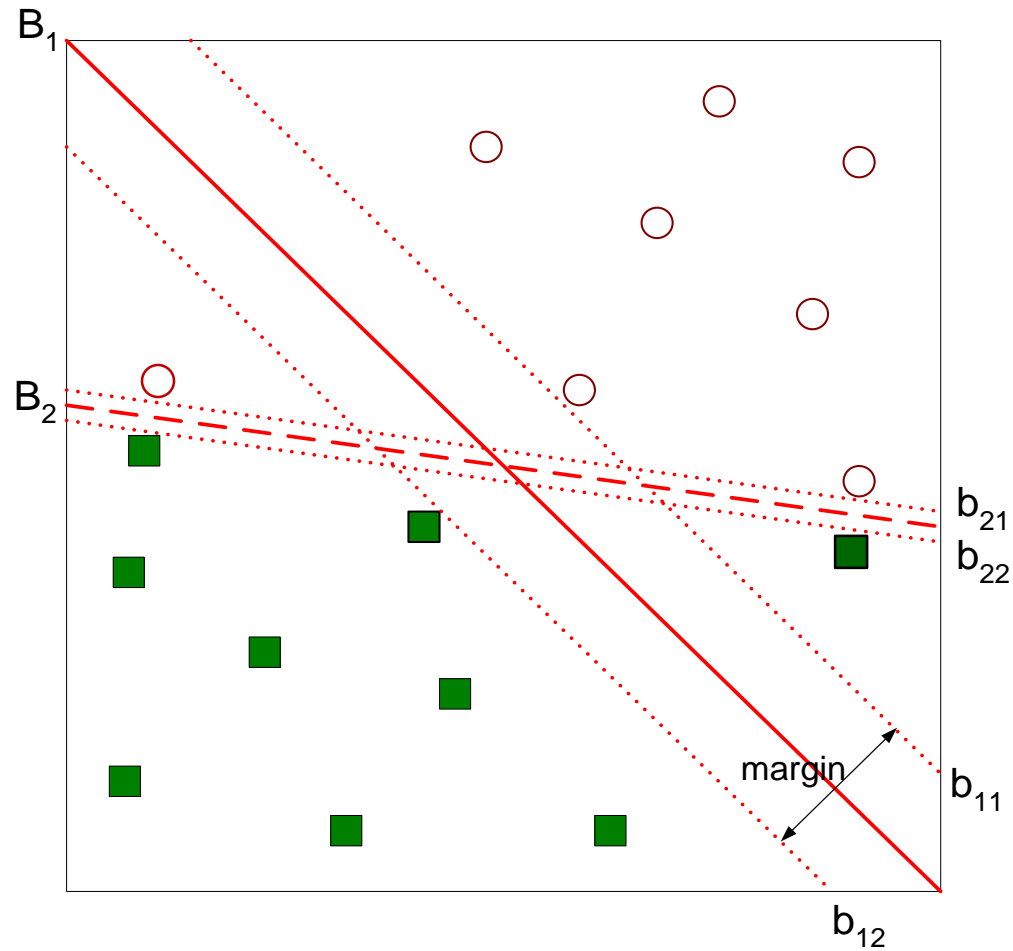
- Precisamos minimizar:

$$L(w) = \frac{\|\vec{w}\|^2}{2} + C \left(\sum_{i=1}^N \xi_i \right)$$

- Sujeito a:

$$y_i = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x}_i + b \geq 1 - \xi_i \\ -1 & \text{if } \vec{w} \bullet \vec{x}_i + b \leq -1 + \xi_i \end{cases}$$

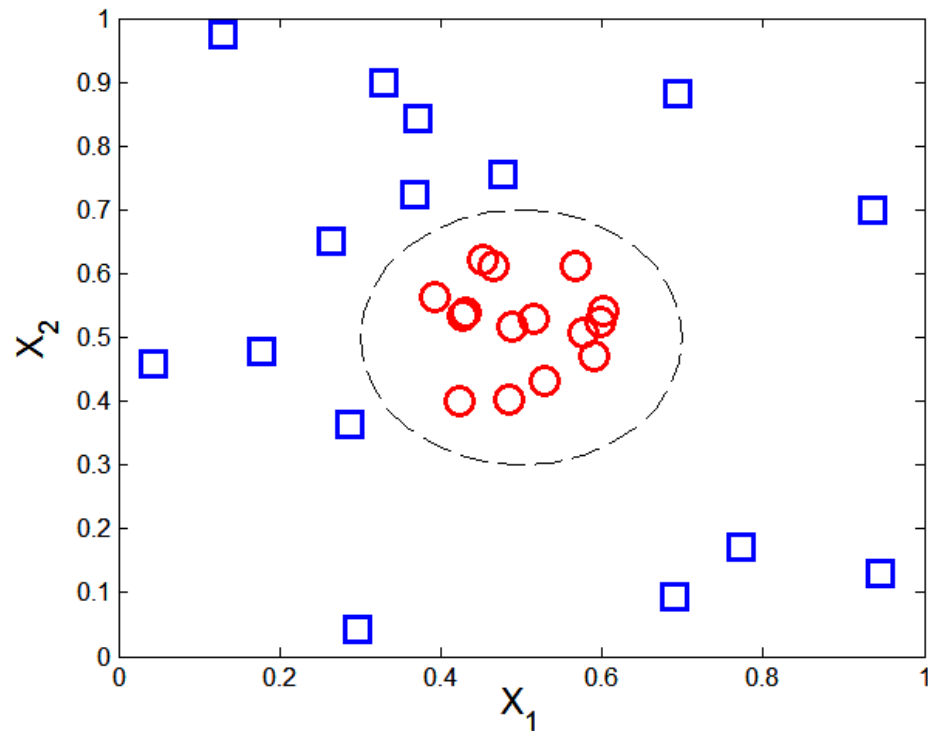
Support Vector Machines



- Encontre o hiperplano que otimiza os dois fatores

Support Vector Machines – caso não linear

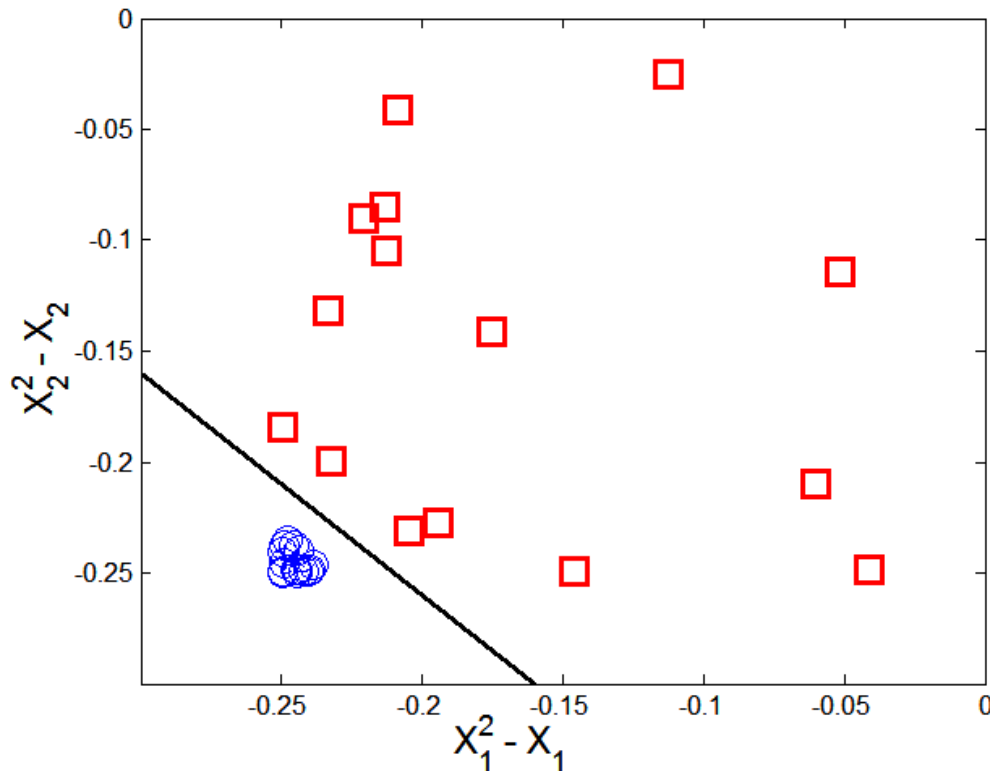
- E se o limite de decisão não for linear?



$$y(x_1, x_2) = \begin{cases} 1 & \text{if } \sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2} > 0.2 \\ -1 & \text{otherwise} \end{cases}$$

Support Vector Machines – caso não linear

- Truque: Transforme dados em um espaço



$$x_1^2 - x_1 + x_2^2 - x_2 = -0.46.$$

$$\Phi : (x_1, x_2) \longrightarrow (x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, 1).$$

$$w_4x_1^2 + w_3x_2^2 + w_2\sqrt{2}x_1 + w_1\sqrt{2}x_2 + w_0 = 0.$$

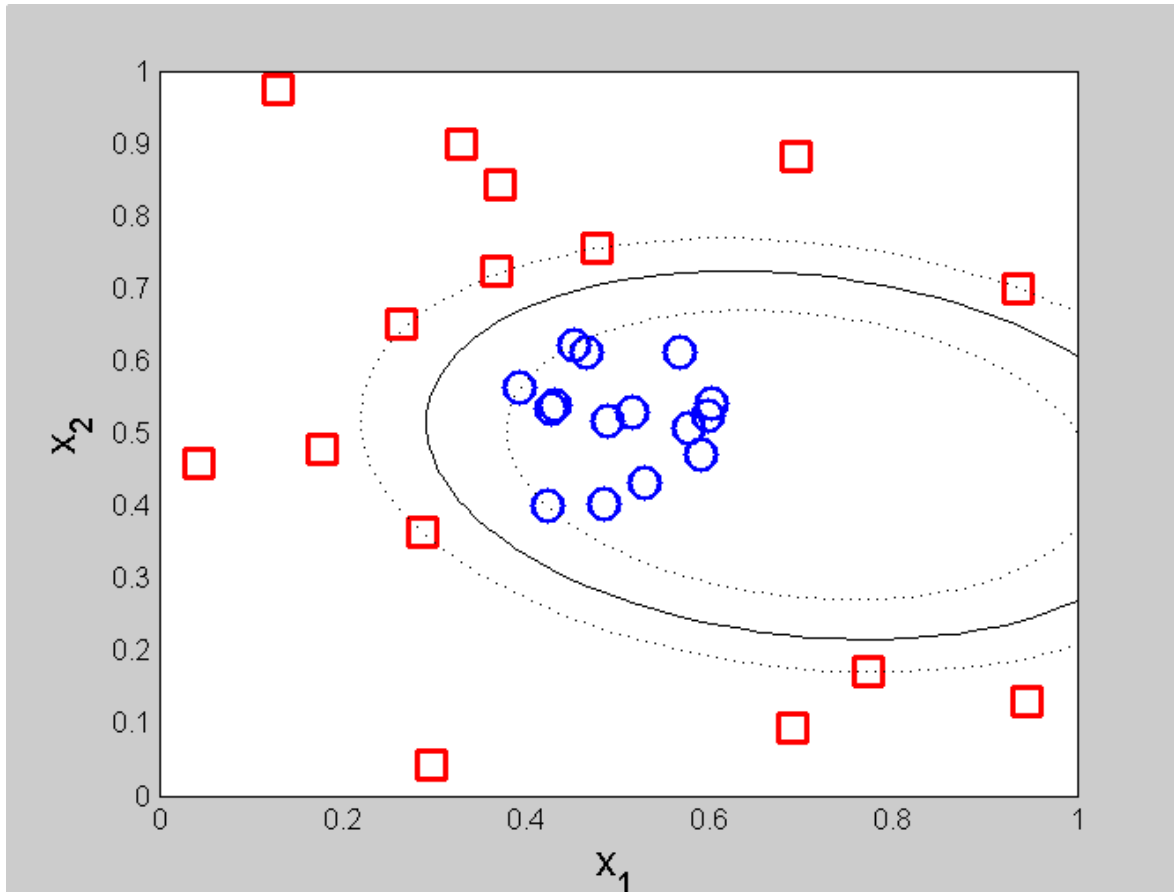
Limite de decisão:

$$\vec{w} \bullet \Phi(\vec{x}) + b = 0$$

Aprendizagem de SVM não linear

- Problemas:
 - Que tipo de função de mapeamento Φ deve ser usado?
 - Como fazer a computação no espaço dimensional elevado?
 - A maioria de computações envolvem o produto escalar
$$\Phi(x_i) \bullet \Phi(x_j)$$
 - Maldição da dimensionalidade?

Exemplo de SVM não linear



**SVM com o kernel
polinômio de grau 2**

Características do SVM

- Uma vez que o problema de aprendizagem é formulado como um problema de otimização convexa, algoritmos eficientes estão disponíveis para encontrar os mínimos globais da função objetiva (muitos dos outros métodos usam abordagens gulosa e encontrar soluções ideais localmente)
- O overfitting é resolvido pela maximização da margem do limite de decisão, mas o usuário ainda precisa fornecer o tipo de função de kernel e função de custo
- Difícil lidar com valores ausentes
- Robusto ao ruído
- Alta complexidade computacional para a construção do modelo

Classificador baseado em regras

- Classifique registros usando uma coleção de regras tipo "se... então... "
- Regra: $(Condição) \rightarrow y$
 - onde
 - *Condição* é uma conjunção de atributos
 - y é o rótulo da classe
 - *LHS*: antecedente da regra ou condição
 - *RHS*: regra conseqüente
 - Exemplos de regras de classificação:
 - $(\text{Blood Type}=\text{Warm}) \wedge (\text{Lay Eggs}=\text{Yes}) \rightarrow \text{Birds}$
 - $(\text{Taxable Income} < 50\text{K}) \wedge (\text{Refund}=\text{Yes}) \rightarrow \text{Evade}=\text{No}$

Classificador baseado em regras (exemplo)

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds
cat	warm	yes	no	no	mammals
leopard shark	cold	yes	no	yes	fishes
turtle	cold	no	no	sometimes	reptiles
penguin	warm	no	no	sometimes	birds
porcupine	warm	yes	no	no	mammals
eel	cold	no	no	yes	fishes
salamander	cold	no	no	sometimes	amphibians
gila monster	cold	no	no	no	reptiles
platypus	warm	no	no	no	mammals
owl	warm	no	yes	no	birds
dolphin	warm	yes	no	yes	mammals
eagle	warm	no	yes	no	birds

R1: (Give Birth = no) \wedge (Can Fly = yes) \rightarrow Birds

R2: (Give Birth = no) \wedge (Live in Water = yes) \rightarrow Fishes

R3: (Give Birth = yes) \wedge (Blood Type = warm) \rightarrow Mammals

R4: (Give Birth = no) \wedge (Can Fly = no) \rightarrow Reptiles

R5: (Live in Water = sometimes) \rightarrow Amphibians

Aplicação de classificador baseado em regras

- Uma regra r **abrange** uma instância x se os atributos da instância satisfizerem a condição da regra

R1: (Give Birth = no) \wedge (Can Fly = yes) \rightarrow Birds

R2: (Give Birth = no) \wedge (Live in Water = yes) \rightarrow Fishes

R3: (Give Birth = yes) \wedge (Blood Type = warm) \rightarrow Mammals

R4: (Give Birth = no) \wedge (Can Fly = no) \rightarrow Reptiles

R5: (Live in Water = sometimes) \rightarrow Amphibians

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
hawk	warm	no	yes	no	?
grizzly bear	warm	yes	no	no	?

A regra R1 cobre um falcão (hawk) \Rightarrow Bird

A regra R3 abrange o urso pardo (grizzly) \Rightarrow Mammal

Cobertura e precisão de regra

- Cobertura de uma regra:
 - A fração de registros que satisfazem o antecedente de uma regra
- Precisão de uma regra:
 - A Fração de registros que satisfazem o antecedente que também satisfaz o consequente de uma regra

<i>Tid</i>	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(Status=Single) → No

Coverage = 40%, Accuracy = 50%

Como funciona o classificador baseado em regras?

R1: (Give Birth = no) \wedge (Can Fly = yes) \rightarrow Birds

R2: (Give Birth = no) \wedge (Live in Water = yes) \rightarrow Fishes

R3: (Give Birth = yes) \wedge (Blood Type = warm) \rightarrow Mammals

R4: (Give Birth = no) \wedge (Can Fly = no) \rightarrow Reptiles

R5: (Live in Water = sometimes) \rightarrow Amphibians

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
lemur	warm	yes	no	no	?
turtle	cold	no	no	sometimes	?
dogfish shark	cold	yes	no	yes	?

Um lemur aciona a regra R3, por isso é classificado como um mamífero.

Uma tartaruga (turtle) encaixa nas regras R4 e R5.

O tubarão dogfish encaixa em nenhuma das regras.

Conjunto ordenado de regras

- As regras são ordenadas de acordo com a sua prioridade
 - Um conjunto de regras ordenada é conhecido como uma lista de decisão
- Quando um registro de teste é apresentado ao classificador
 - Ele é atribuído ao rótulo de classe da regra de classificação mais alta que disparou
 - Se nenhuma das regras disparar, ele é atribuído à classe padrão

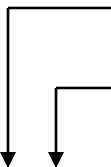
R1: (Give Birth = no) \wedge (Can Fly = yes) \rightarrow Birds

R2: (Give Birth = no) \wedge (Live in Water = yes) \rightarrow Fishes

R3: (Give Birth = yes) \wedge (Blood Type = warm) \rightarrow Mammals

R4: (Give Birth = no) \wedge (Can Fly = no) \rightarrow Reptiles

R5: (Live in Water = sometimes) \rightarrow Amphibians



Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
turtle	cold	no	no	sometimes	?

Esquemas de ordenação de regras

- Ordenação baseada em regras
 - As regras individuais são classificadas com base na sua qualidade
- Ordenação baseada em classe
 - As regras que pertencem à mesma classe aparecem juntas

Rule-based Ordering

(Refund=Yes) ==> No

(Refund=No, Marital Status={Single,Divorced}, Taxable Income<80K) ==> No

(Refund=No, Marital Status={Single,Divorced}, Taxable Income>80K) ==> Yes

(Refund=No, Marital Status={Married}) ==> No

Class-based Ordering

(Refund=Yes) ==> No

(Refund=No, Marital Status={Single,Divorced}, Taxable Income<80K) ==> No

(Refund=No, Marital Status={Married}) ==> No

(Refund=No, Marital Status={Single,Divorced}, Taxable Income>80K) ==> Yes

Vantagens de classificadores de regras

- Tem características bastante semelhantes às árvores de decisão
 - altamente expressivos como árvores de decisão
 - Fácil de interpretar
 - Desempenho comparável a árvores de decisão
 - Pode manipular atributos redundantes
- Melhor adequados para lidar com classes desequilibradas
- Mais difícil lidar com valores ausentes no conjunto de testes

Métodos baseados nas assembleias

- Construir um conjunto de classificadores dos dados de treinamento
- Prever o rótulo de classe de registros de teste combinando as previsões feitas por vários classificadores

Abordagem geral

