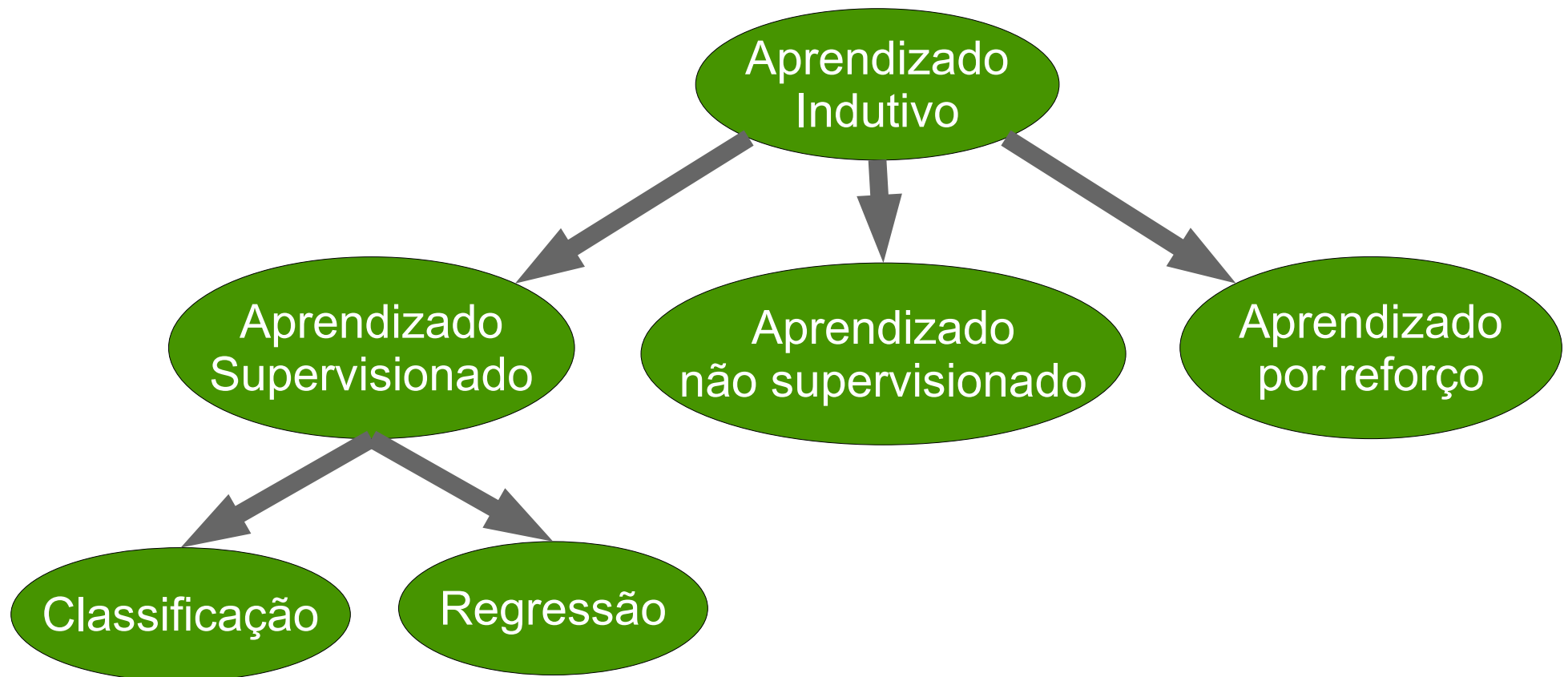


Aprendizado de Máquina: Classificação de Dados

Prof. Arnaldo Candido Junior
UTFPR – Medianeira

Abordagens de Aprendizado



Classificação

- Atribuir objetos a uma dentre várias categorias pré-definidas
- Ex.:
 - Classificação de letras e números
 - Reconhecimento de faces
 - Análise de crédito
 - Diagnóstico médico

Classificação ₍₂₎

- Dado um conjunto de treinamento, em que cada exemplo possui um conjunto de atributos e um deles o rótulo ou classe
 - Encontrar um modelo para o atributo classe como uma função dos valores de entrada
 - Função alvo ou modelo de classificação

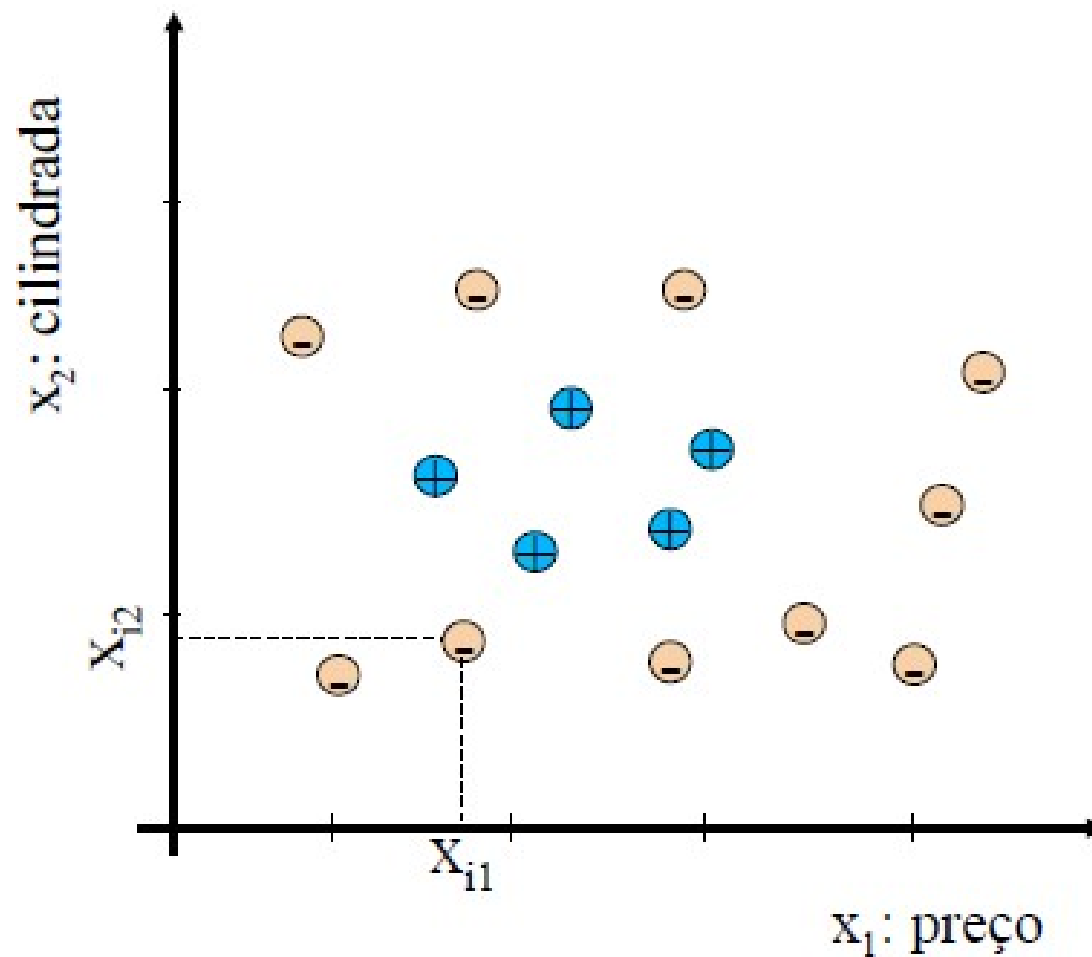
Classificação₍₃₎

- Utilidades do modelo de classificação
 - **Modelagem descritiva**: explica como são discriminados objetos de diferentes classes
 - Extração de conhecimento: resume os dados
 - **Modelagem preditiva**: permite prever o rótulo de novos dados de entrada
 - Não vistos anteriormente

Classificação ₍₄₎

- Supor a tarefa de aprender a classificar carros em duas classes
 - Carro esporte (+)
 - Carro passeio (-)
- Dados de entrada:
 - Características de um carro
 - Preço (x_{i1}) e cilindrada (x_{i2})

Classificação ₍₅₎



Classificação ₍₆₎

- Dados exemplos de treinamento, encontrar um modelo
- **Modelagem descritiva**: o que representa um carro de passeio?
- **Modelagem preditiva**: qual a classe de um novo carro?

Teoria do Aprendizado Estatístico

- Fornece uma interpretação estatística sobre o aprendizado de máquina
- Vamos usar uma notação diferente da tradicional
 - Notações variam muito de acordo com autor ou algoritmo

Notação

- x, x_1, u, v, \dots : valores de instâncias (sem a classe) ou escalares
- $\hat{x}, \hat{x}_1, \hat{u}, \hat{v}, \dots$: instâncias ou pontos ou vetores
Note que: $\hat{x} = (x_1, x_1, \dots, x_n)$
- $f(\hat{x}), d$: **classe real** (ou desejada) de uma instância.
 - Vale 1 ou 0 na classificação binária (foco inicial) (alternativamente, “+” ou “-”)

Notação ₍₂₎

- h : hipótese – o recorte no espaço que permite que a classificação aconteça
- $h(\hat{x})$, y : **classe predita** para uma instância
- h^* : hipótese ótima – aquela que mais se aproxima de f
- X : um dataset ou uma matrizes

Notação ₍₃₎

- Exemplo:
 - Instância Gol (\hat{x});
 - Preço: R\$ 30.000 (x_1)
 - Cilindrada: 1.000 (x_2)
 - Classe: passeio (0 – classe positiva)
 - Notação:
 $\hat{x} = (30.000, 1.000)$
 $f(\hat{x}) = 0$

Notação ₍₄₎

- Uma instância é dada por seus atributos:
 $\hat{x} = (x_1, x_2, \dots, x_n)$
- Um conjunto de dados (dataset) é dado por várias instâncias

$$X = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & & & x_{2n} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

Notação ₍₅₎

- No aprendizado supervisionado, os datasets são rotulados, isto é, dados são pares \hat{x} e y

$$X = \begin{bmatrix} \hat{x}_1 & d_1 \\ \hat{x}_2 & d_2 \\ \vdots & \vdots \\ \hat{x}_n & d_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} & d_1 \\ x_{21} & & & x_{2n} & d_2 \\ \vdots & & & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} & d_n \end{bmatrix}$$

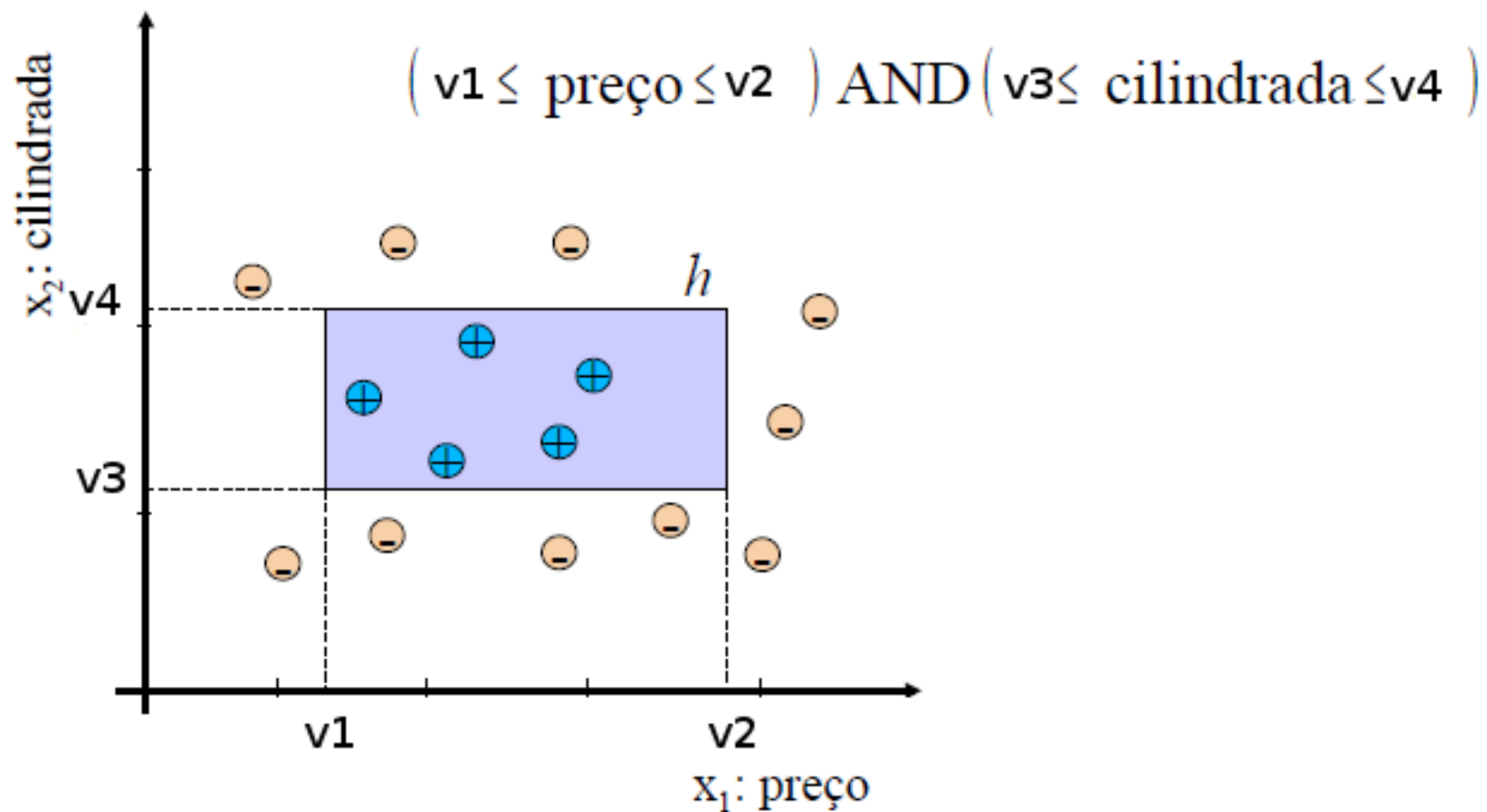
Teoria do Aprendizado Estatístico ⁽²⁾

- Como exemplo, utilizaremos hipóteses na forma de retângulos (hipercubos)
 - São mais fáceis de interpretar
 - Formato:
$$h(\hat{x}) = \begin{cases} 1 & \text{Se } v_1 \leq x_1 \leq v_2 \text{ E } v_3 \leq x_2 \leq v_4 \\ 0 & \text{caso contrário} \end{cases}$$
 - v_1, v_2, v_3 e v_4 delimitam o hipercubo

Teoria do Aprendizado Estatístico ⁽³⁾

- Exemplo:
 - $h(\hat{x}) = 0$ **Se** $1.000 \leq x_1 \leq 30.000$ **E**
 $1.000 \leq x_2 \leq 2.000$
1 caso contrário

Teoria do Aprendizado Estatístico ⁽⁴⁾



Teoria do Aprendizado Estatístico (5)

- **Falso positivo** (classe de referência: passeio):
 $h(\hat{x}) = 1; f(\hat{x}) = 0$
- **Falso negativo**:
 $h(\hat{x}) = 0; f(\hat{x}) = 1$
- **Erro empírico** conta o número de erros que a hipótese h comete em um determinado dataset D

$$e(h|X) = \sum_{\hat{x} \in X} h(\hat{x}) \neq f(\hat{x})$$

Teoria do Aprendizado Estatístico (6)

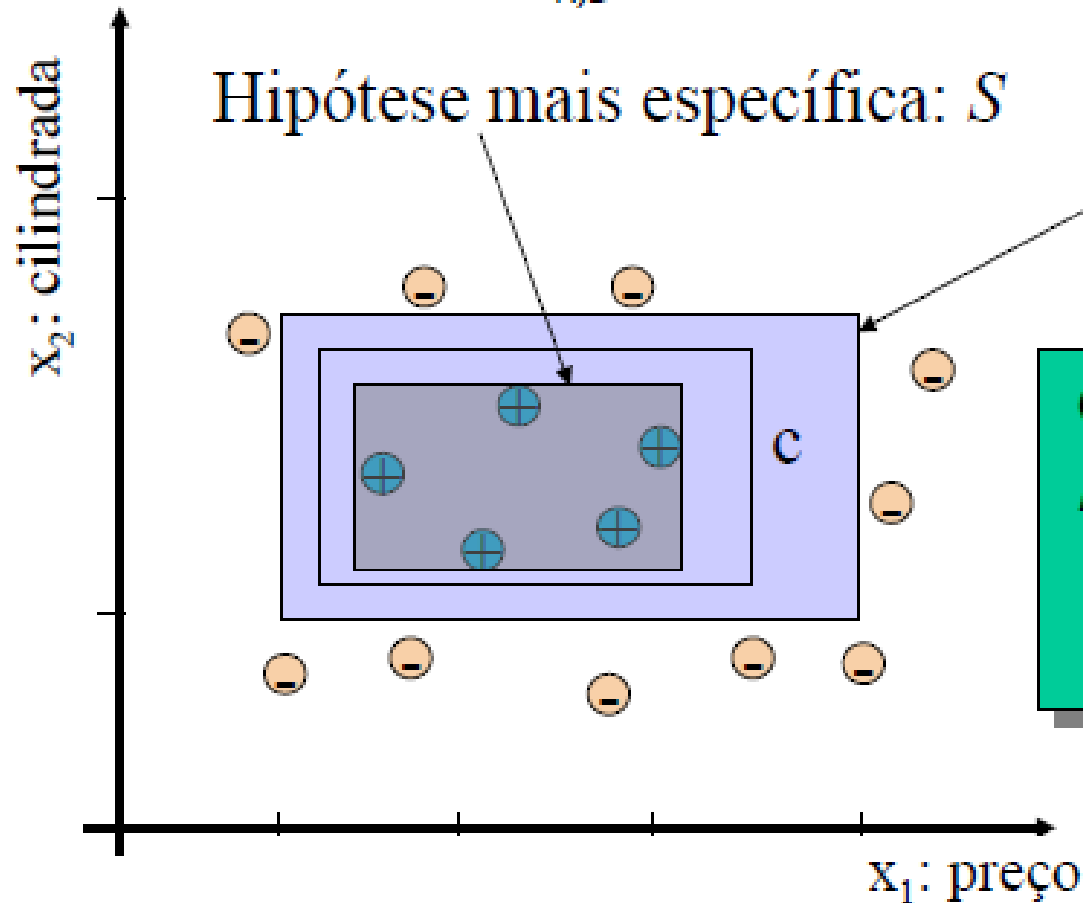
- Chamaremos de H o conjunto de todas as hipóteses possíveis sobre um dataset
 - Normalmente, H é infinito
- Tentaremos encontrar h que melhor **generalize** os dados (isto é, h^*)
- h^* idealmente deve incluir todos os exemplos positivos e nenhum exemplo negativo
- Obs: pode haver mais de um h^*

Teoria do Aprendizado Estatístico ⁽⁷⁾

$$EV_{H,D} \equiv \{h \in H \mid \text{Consistente}(h, X)\}$$

Hipótese mais específica: S

Hipótese mais geral: G



Qualquer hipótese $h \in H \mid S \leq h \leq G$:
É válida e consistente com D
Faz parte do EV

Teoria do Aprendizado Estatístico ⁽⁸⁾

- **Fronteira específica:**
 - É mais propensa a falsos negativos para dados não vistos em treinamento
- **Fronteira geral:**
 - É mais propensa a falsos positivos

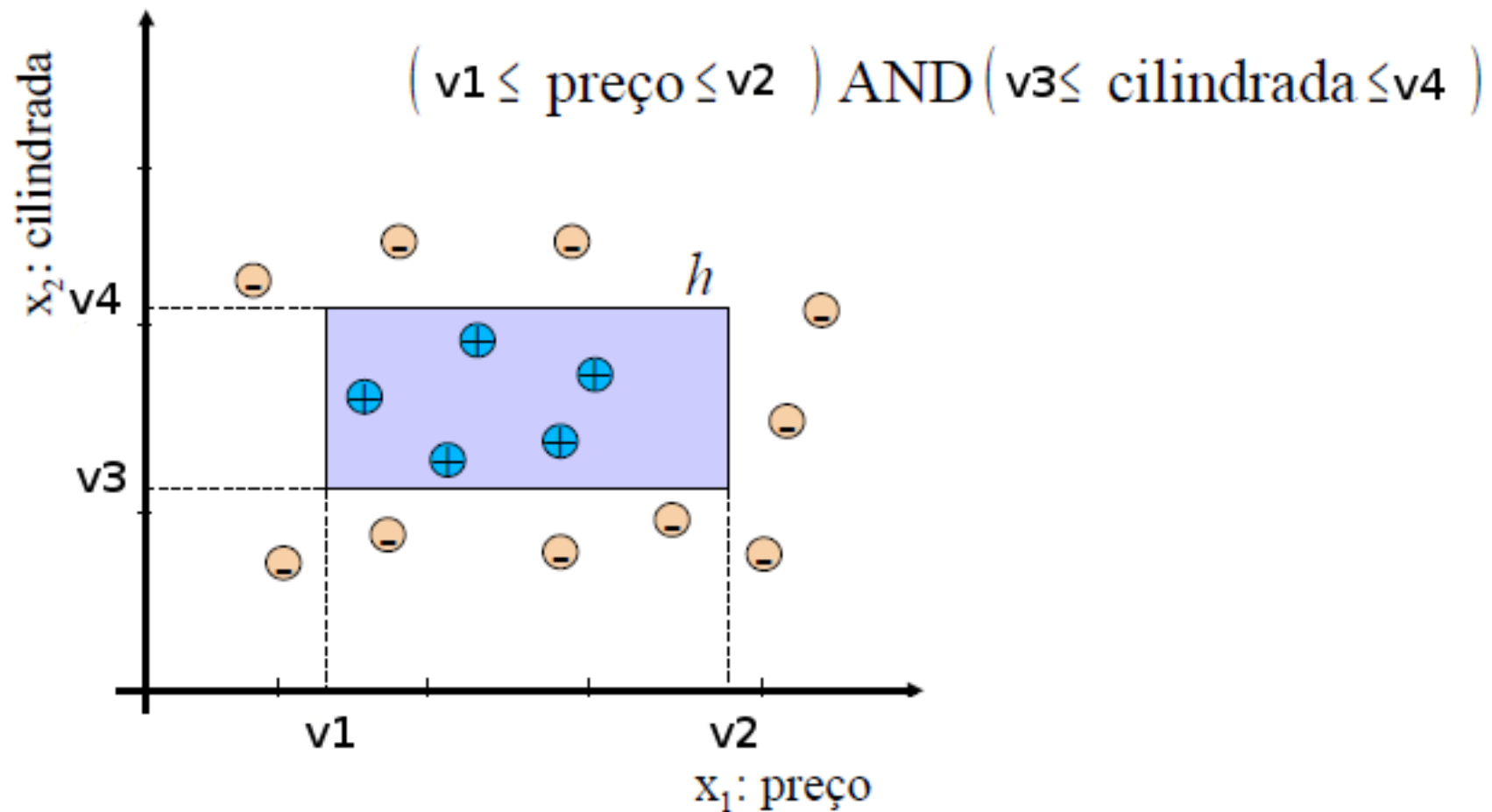
Espaços

- Temos dois espaços importantes
 - **Espaço de instâncias**
 - **Espaço de hipóteses**
 - Subespaço de versões

Espaço de instâncias

- Cada ponto é uma instância
- Tem a forma de nuvem de pontos
- Objetivo: recortar esse espaço
- É o que estudamos até o momento:
 - No exemplo dado, é o R^2

Espaço de instâncias ⁽²⁾



Espaço de hipóteses

- Cada ponto representa uma hipótese / recorte / classificador
- Tem a forma de encosta
- Objetivo: descer a encosta

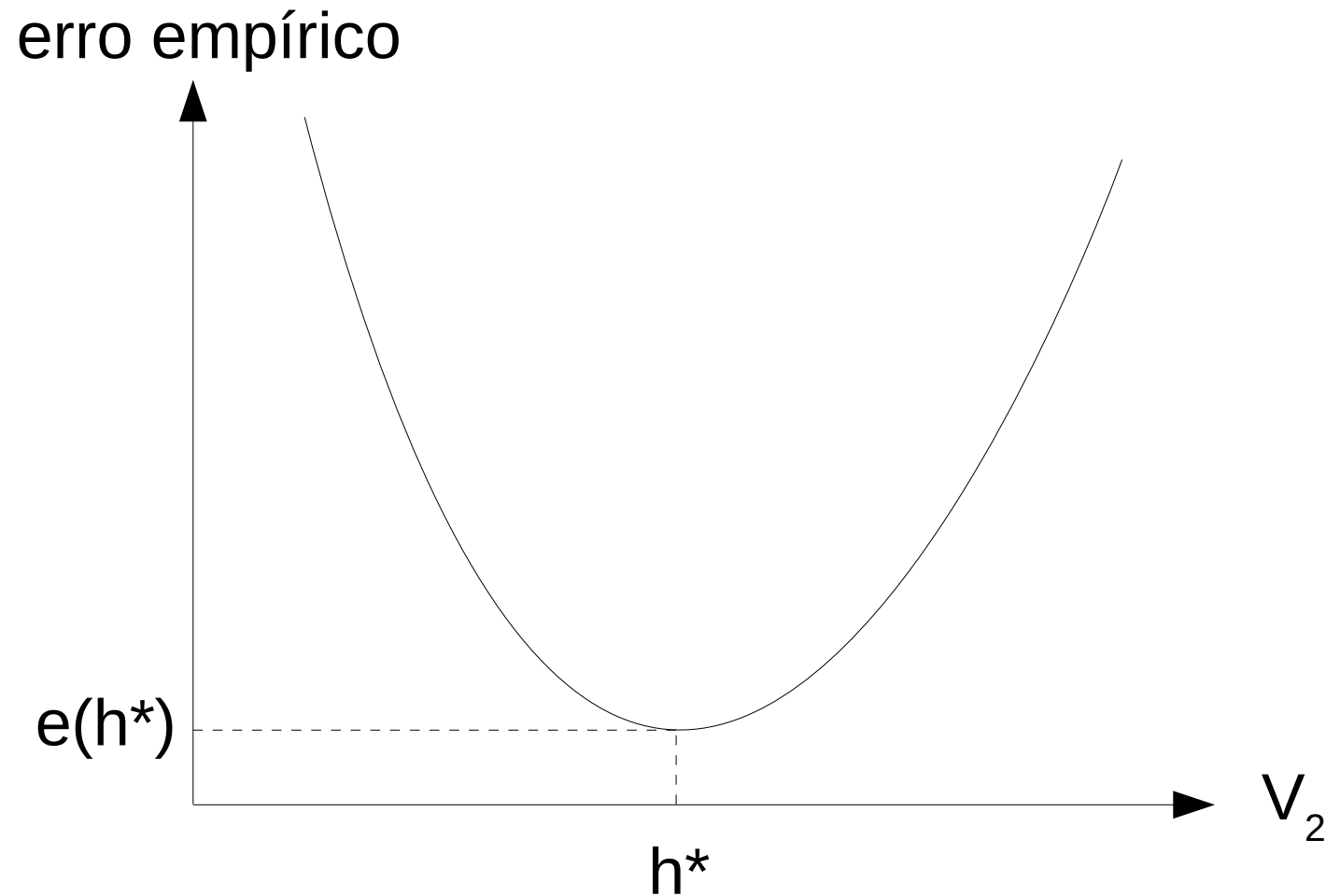
Espaço de hipóteses

- Todo o recorte no espaço de instâncias se transforma em um ponto no espaço de hipóteses
- Precisamos converter nossos retângulos e seus respectivos erros empíricos em pontos
 - No exemplo, isto pode ser feito da seguinte forma: $h = (v_1, v_2, v_3, v_4, e)$

Espaço de hipóteses ₍₂₎

- Versão original (R^5): $h = (v_1, v_2, v_3, v_4, e)$
- Simplificando para visualização (R^2): $H = (v_2, e)$
 - Lembrando: v_2 armazena o preço máximo de um veículo de passeio
 - v_2 grande \rightarrow falsos negativos
 v_2 pequeno \rightarrow falsos positivos

Espaço de hipóteses ₍₃₎



Espaço de hipóteses (4)

- Regra informal: coisas parecidas andam juntas
 - h 's parecidos possuem erros empíricos parecidos
 - Percorremos hipóteses em H em busca de h^*
 - Formalmente: estamos minimizando o erro empírico

Espaço de hipóteses (5)

- Classificadores fazem uma **busca** no espaço de hipóteses
 - **Informada**: no exemplo, heurística = erro empírico
 - Normalmente **gulosa**: se não encontrar h^* , se contenta com uma aproximação boa o bastante h'
 - Normalmente **descida de encosta** (down hill): minimiza uma função de custo
 - Obs: erro empírico não é a melhor função de custo, pois não é suave, existem outras...

Subespaço de versões

- É um pedaço do espaço de hipóteses no qual...
 - ... cada hipótese separa **perfeitamente** o conjunto de treinamento X (erro empírico = 0)
 - Ou seja, ótimo global da encosta
 - Pode ser vazio (comum)
 - Pode ter uma ou mais fronteiras específicas
 - Pode ter uma ou mais fronteiras gerais

Subespaço de versões ₍₂₎

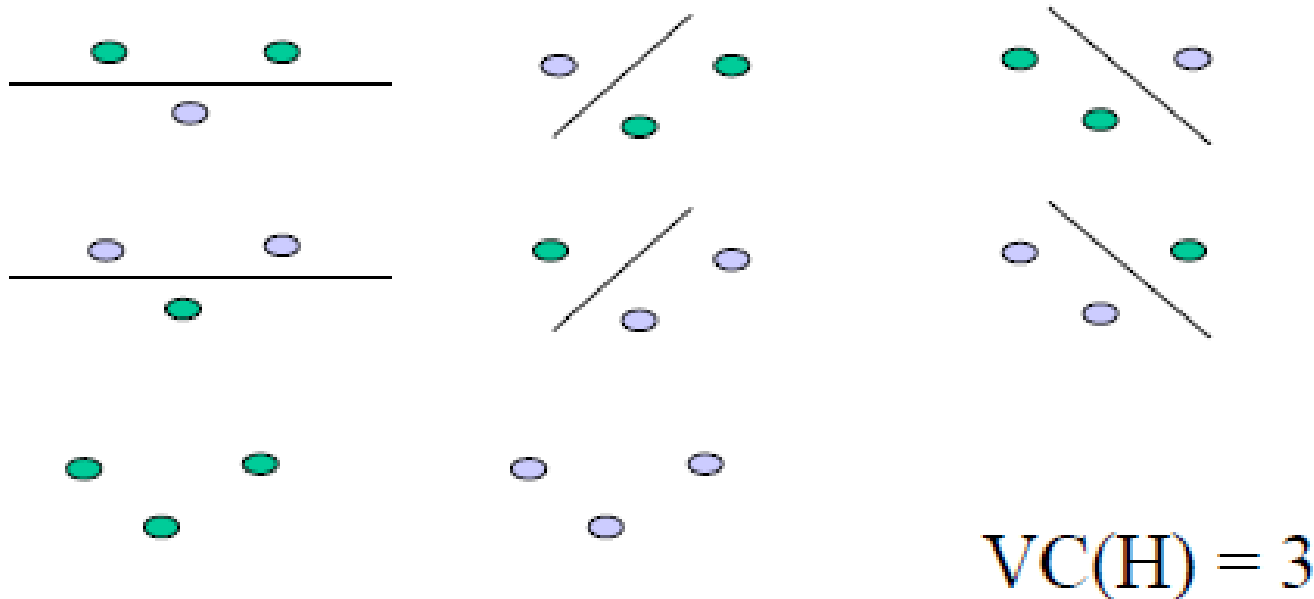
- É propenso a hipóteses que fazem **overfitting**
 - Acertar o **conjunto de treinamento** sugere memorização
 - Por causa disso, avaliações normalmente são feitas por um dataset separado, chamado de **conjunto de teste**

Dimensão VC

- Dá uma ideia do poder de um classificador
 - E também se ele é propenso a overfitting
- N pontos podem ser rotulados de 2^N possíveis maneiras como +/-
- Mede a capacidade de um classificador de classificar corretamente esses 2^N problemas
 - Dimensão varia de acordo com conjunto de treinamento

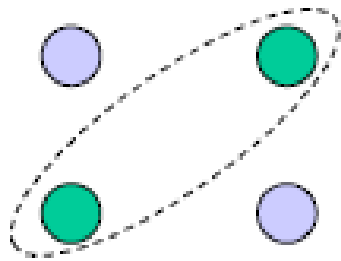
Dimensão VC₍₂₎

- Seja um classificador linear em um espaço bidimensional
 - Para os três pontos abaixo, temos 8 problemas de separação (2^3)
 - Todos os problemas são linearmente separáveis



Dimensão VC₍₃₎

- Seja um classificador linear em um espaço bidimensional
- $N = 4$ -- 16 problemas
 - Alguns problemas não são linearmente separáveis

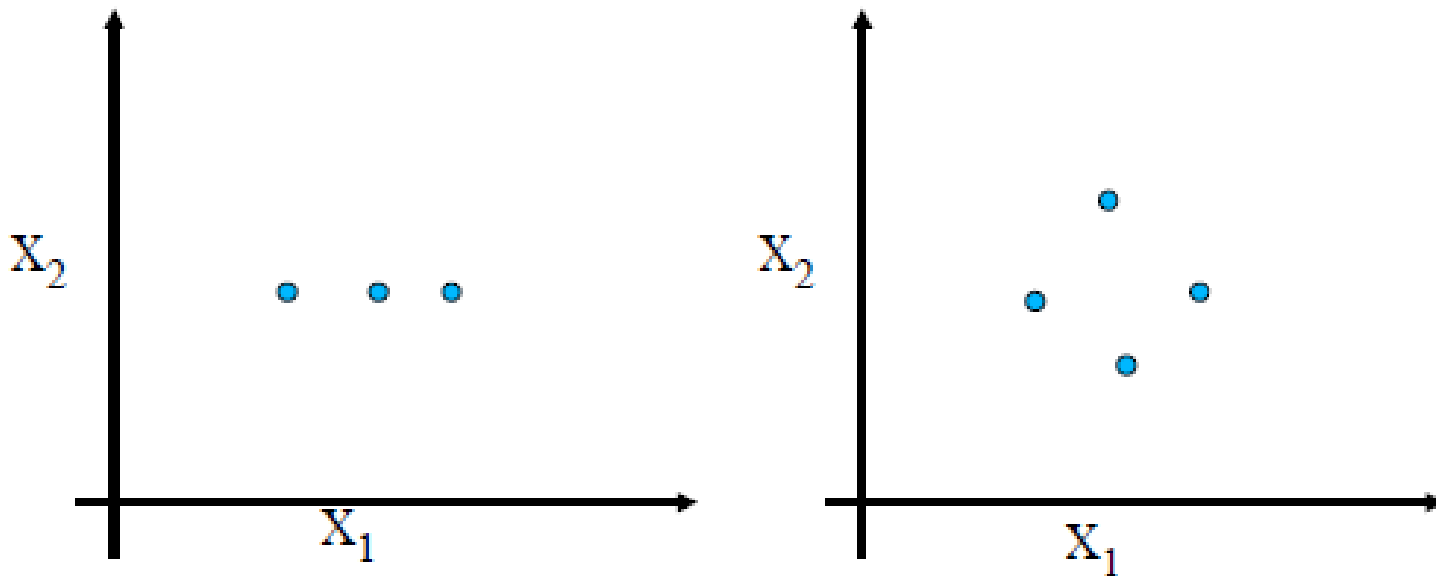


$$VC(H) = 3$$

$VC(H)$ para um classificador linearmente separável em um espaço bidimensional é 3

Dimensão VC₍₄₎

- Depende de:
 - Espaço de hipóteses
 - Número de atributos
 - Quais são os N dados escolhidos de X

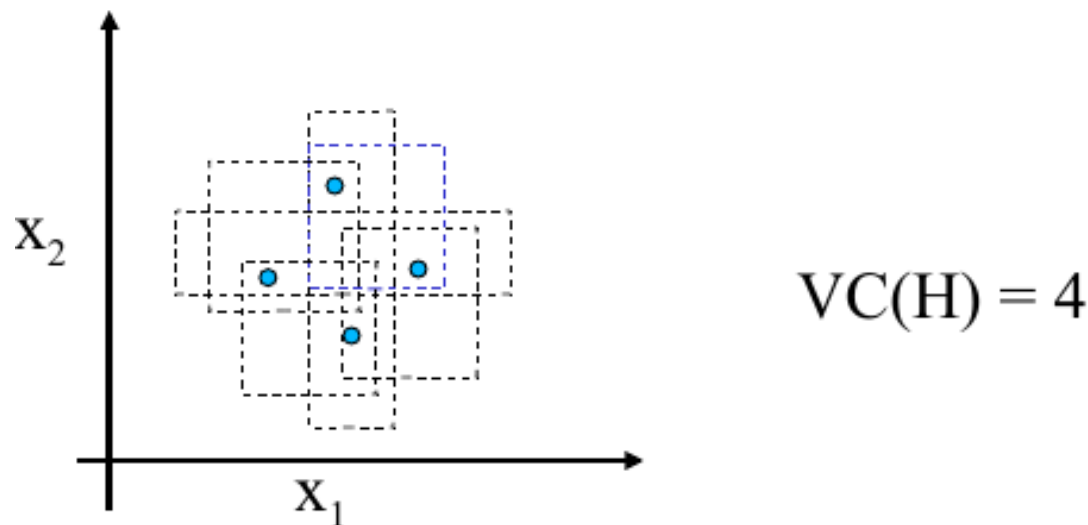


Dimensão VC ₍₅₎

- Seja um classificador na forma de retângulo em um espaço bidimensional
 - $N = 4$ (16 problemas)

Dimensão VC ₍₆₎

- Seja um classificador na forma de retângulo em um espaço bidimensional
 - $N = 4$ (16 problemas)
 - Todos os problemas são separáveis por um retângulo



Dimensão VC ₍₇₎

- Independe da distribuição de probabilidade com que exemplos são gerados
 - Depende do número de atributos
 - Depende da representação das hipóteses
 - Se H for uma tabela de busca
 - Dimensão VC é infinita

Dimensão VC ₍₈₎

- Pode parecer pessimista
 - Muitas das possíveis combinações dado-classe não ocorrem na prática
 - Dados reais são mais regulares: exemplos de uma mesma classe tendem a estar próximos
 - “Coisas parecidas andam juntas”
- Na prática, classificadores separam conjuntos maiores de dados

Dimensão VC ₍₉₎

- Dimensão VC é um conceito teórico
 - É difícil de calcular com exatidão para a maioria dos classificadores
 - Quanto mais flexível o classificador, maior tende a ser sua dimensão VC

Classificação

- Técnica de classificação: abordagem sistemática para construir um classificador
 - Ex. RNs, SVMs, árvores de decisão
 - Cada técnica emprega um algoritmo de aprendizado para procurar pela melhor hipótese no espaço de hipóteses
 - Modelo gerado deve se ajustar aos dados e permitir a predição da classe de novos exemplos

Classificação ₍₂₎

conjunto de dados

x_{11}	x_{12}	...	x_{1m}	d_1
x_{21}	x_{22}	...	x_{2m}	d_2
\vdots	\vdots		\vdots	\vdots
x_{n1}	x_{n2}	...	x_{nm}	d_n

Técnica de AM



$h(x)$

Classificador

Treinamento

\mathbf{x}_i

x_{i1}	x_{i2}	...	x_{im}
----------	----------	-----	----------

novo dado



$h(x)$

Classificador



$h(x)$

previsão

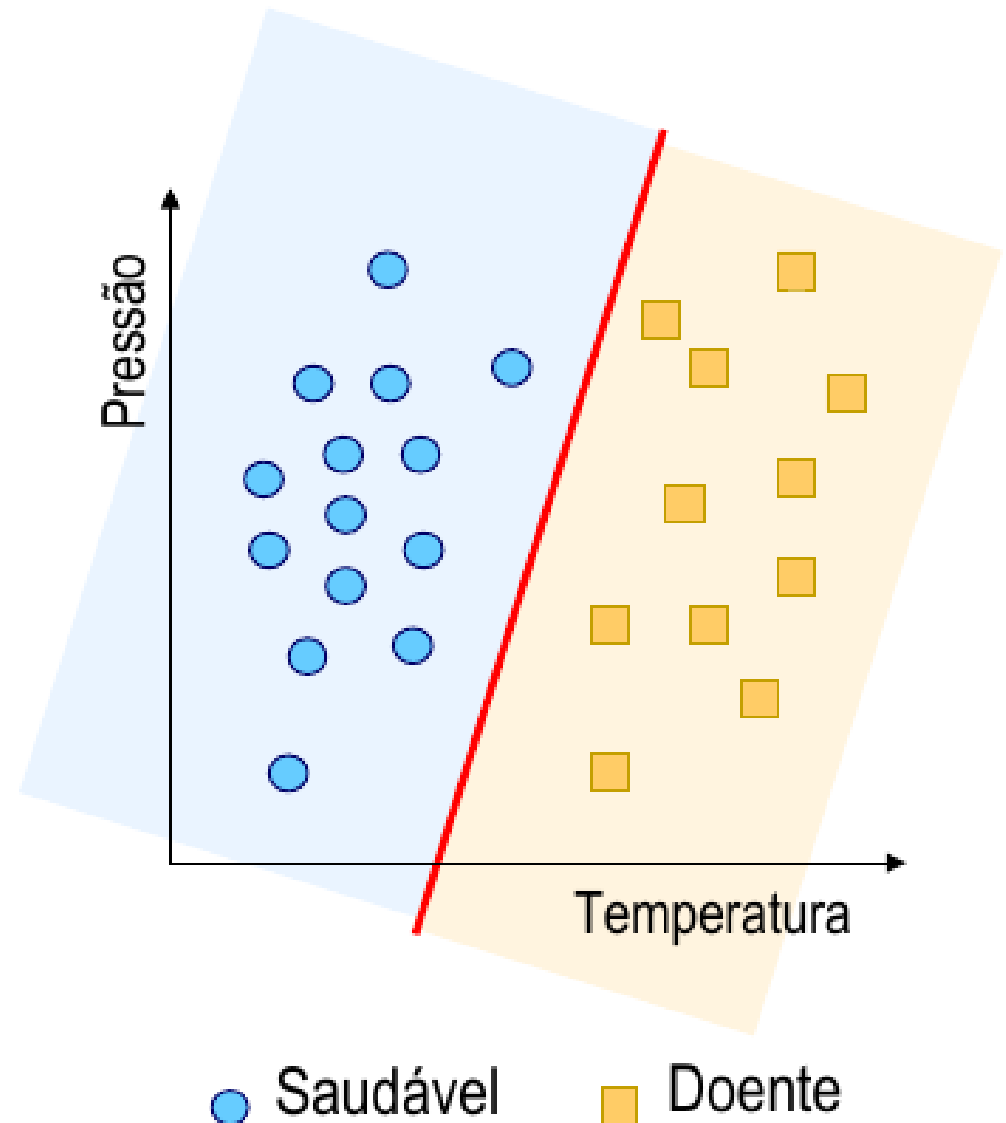
Teste

Tipos de Classificação

- Classificação Binária
- Classificação com uma Classe (detecção de novidades)
- Classificação Multiclasse
- Classificação Multi-rótulos
- Classificação com Ranking
- Classificação Hierárquica
- Existem outras

Classificação Binária

- Mais comum
 - Dados podem pertencer a uma dentre 2 classes
 - Classe positiva
 - Classe negativa

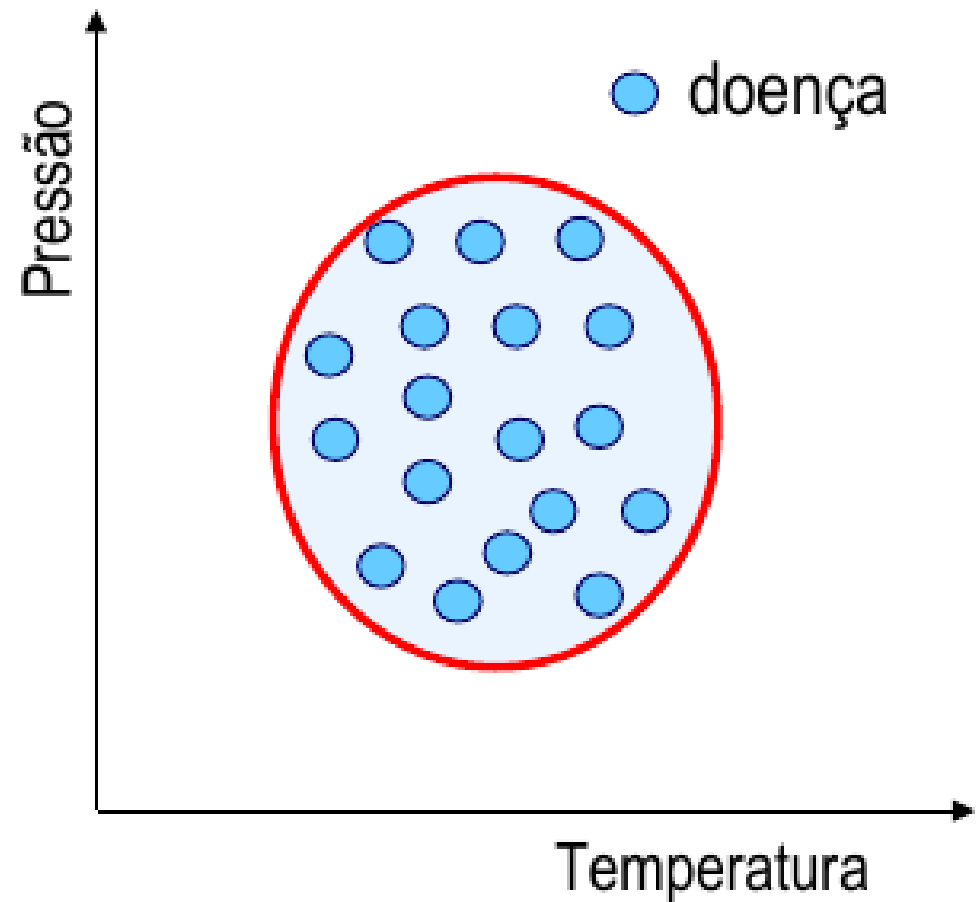


Classificação Binária ₍₂₎

- Existem várias métricas para avaliação dos classificadores
- Problema:
 - Desbalanceamento das classes
 - Nem toda tarefa de classificação é binária

Classificação com Uma Classe

- Dados somente da classe +1 (positiva)
- Ou poucos dados
- Obter exemplos negativos: difícil, custoso



Classificação com Uma Classe ⁽²⁾

- Aprendizado com apenas exemplos positivos
 - Contra-exemplos (normalmente condições anômalas) são caros, difíceis de obter
 - Ou difíceis de caracterizar
- Exemplos negativos são importantes na definição do grau de generalização
 - Ausência dificulta indução de um classificador adequado

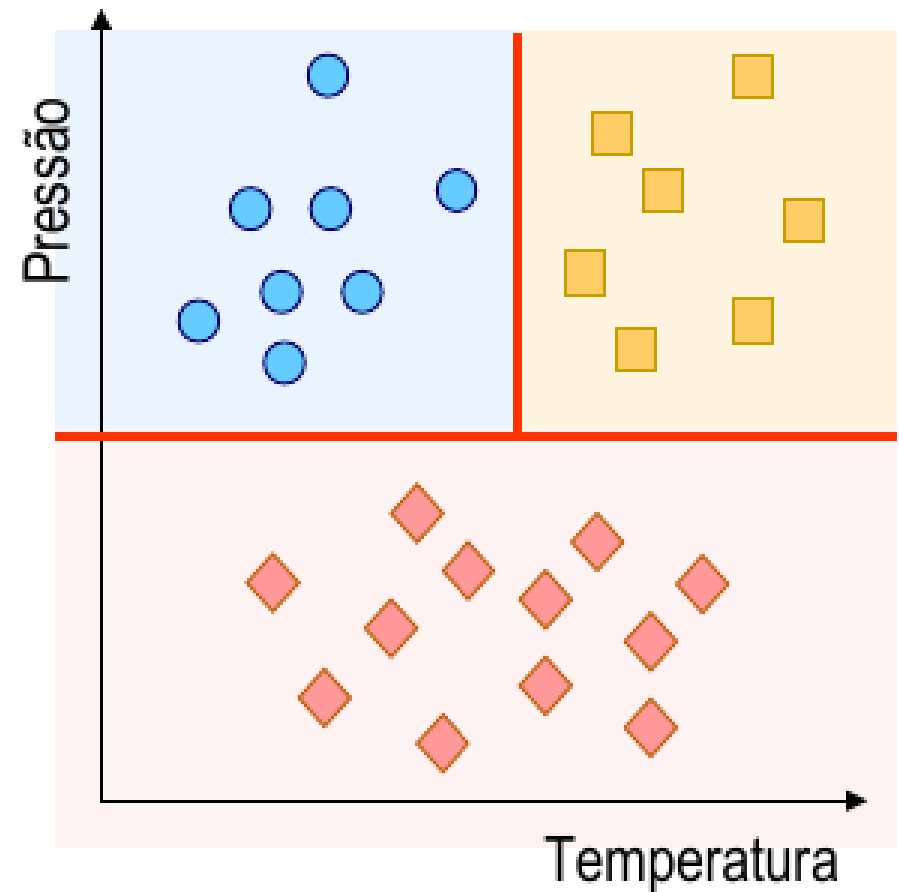
Problemas Multiclasses

- Várias técnicas de AM podem induzir apenas classificadores binários
- Existe um grande número de problemas reais com mais que 2 classes
 - Problemas de classificação multiclasses
 - Para serem resolvidos, são utilizadas estratégias multiclasses

Problemas Multiclasses (2)

- Dados em várias classes
 - $\{1, 2, \dots, k\}$
 - Classes mutuamente excludentes

● gripe
■ sarampo
◆ amigdalite



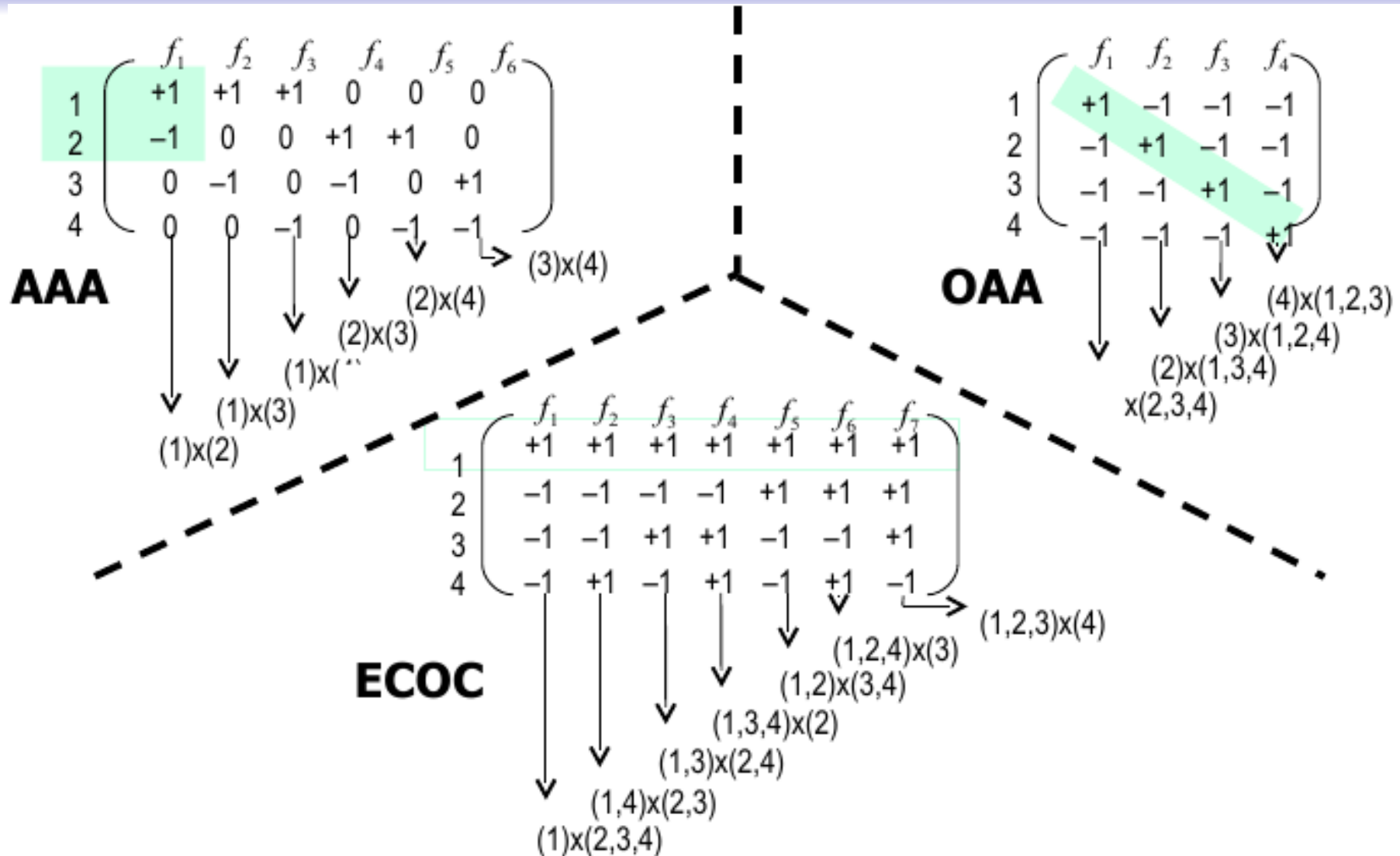
Problemas Multiclasses ⁽³⁾

- Duas abordagens têm sido utilizadas:
 - **Modificação do Algoritmo de classificação**
 - Operações internas refeitas
 - **Decomposição do problema multiclasse**
(foco da aula)
 - Um problema multiclasse é transformado em em múltiplos problemas binários

Problemas Multiclasses

- **Decomposição**
 - Geralmente reduz a complexidade do problema
 - Permite processamento paralelo
 - Veremos decomposição por **Matrizes de Códigos** (a seguir)

Matrizes de códigos



Matrizes de códigos ₍₂₎

- **OAA (one against all):** um contra todos
 - n classes $\rightarrow n$ classificadores
 - C_i : compare a classe i contra todas as demais
 - Classe escolhida: aquela que obteve o maior score

Matrizes de códigos ₍₃₎

- **AAA (all against all)**: todos contra todos
 - C_1 : classe 1 contra classe 2; C_2 : classe 1 contra classe 3; ...
 - Classe escolhida: é feita uma votação em cada classificador
 - Também chamado de “one against one”

Matrizes de códigos ⁽⁴⁾

- **ECOC** (error correction output codes): correção de erro dos códigos de saída
 - Flexibiliza o número de classificadores
 - Dois casos:
 - **Caso 1**: mais classificadores do que classes:
 - Saída mais confiável
 - **Caso 2**: menos classificadores do que classes
 - Menor custo de classificação

Matrizes de códigos ₍₅₎

- **Caso 1:** mais classificadores do que classes
 - Representamos classes por mapas de bits.
 - Para 4 classes, precisamos de 4 bits
 - Projetista define número de classificadores a ser utilizado
 - Por sorteio, cada classificador emite um mapa de bits diferente

Matrizes de códigos ₍₆₎

- **Caso 1:** exemplos
 - Classificador 1: sorteado classe 1 contra o resto
 - Saídas: 1000 para classe positiva, 0111 para negativa
 - Classificador 2: sorteado classes 2 e 3 contra o resto
 - Saídas: 0110 para positiva; 1001 para negativa
 - ...
 - Interpretando a saída: fazer a média ponderada entre os bits de todas as saídas: classe mais votada ganha

Matrizes de códigos ₍₇₎

- **Caso 2:** menos classificadores do que classes
 - Para 4 classes (C_1, C_2, C_3, C_4) e 2 classificadores usamos dois bits: $C_1 = 00$; $C_2 = 01$; $C_3 = 10$; $C_4 = 11$
 - Classificador 1: aprende a emitir primeiro bit
 - $0 = C_1$ ou C_2
 - $1 = C_3$ ou C_4
 - Classificador 2: aprende a emitir segundo bit
 - $0 = C_1$ ou C_3 ;
 - $1 = C_2$ ou C_4

Matrizes de códigos ₍₈₎

- **Exercício:** Considere que a saída ECOC do exemplo anterior foi:

$$f_1 = -1$$

$$f_2 = +1$$

$$f_3 = -1$$

$$f_4 = +1$$

$$f_5 = -1$$

$$f_6 = +1$$

$$f_7 = -1$$

- Calcule a classe escolhida

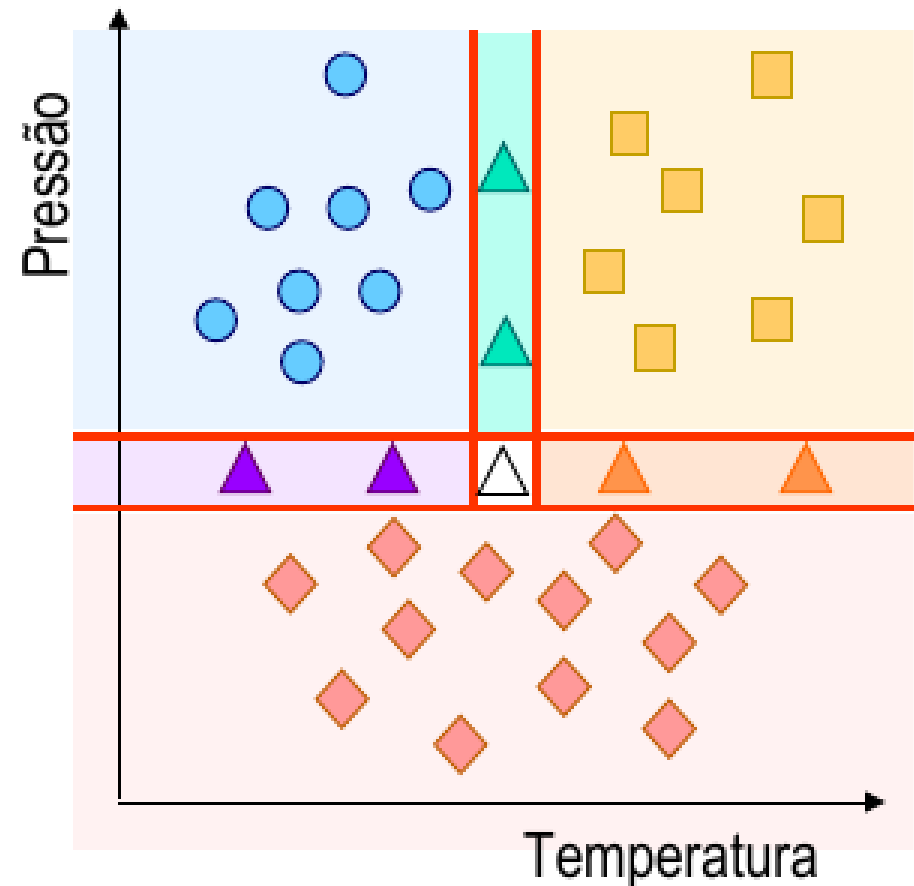
Matrizes de códigos ₍₈₎

- Testar no Weka:
 - Meta-classificador multi-classe para classificadores baseados em duas classes
 - `weka.classifiers.meta.multiclassifier`
 - Testar: AAA, OAO, ECOC

Classificação Multirrótulos

- Dados em várias classes
 - $\{1, 2, \dots, k\}$
 - Classes sobrepostas

● gripe ▲ gripe e sarampo
■ sarampo ▲ gripe e amigdalite
◆ amigdalite ▲ sarampo e amigdalite
△ gripe, sarampo e amigdalite



Classificação Multirrótulos ⁽²⁾

- Duas abordagens:
 - Adaptar algoritmo de classificação para saída multirrótulo
 - Reduzir problema para multiclasse (**foco da aula**)

Classificação Multirrótulos ⁽³⁾

- Estratégias para reduzir problema para multiclasse:
 - 1. Usar matrizes de código
 - 2. Criar rótulos
 - 3. Eliminar instâncias
 - 4. Eliminar rótulos

Classificação Multirrótulos (4)

- **Estratégia 1:** usar matrizes de códigos
 - Reduz multiclasse para binário **OU**
 - Reduz multirrótulo para binário

Problema Multi-rótulo:

Instância 1: classes A and B
Instância 2: classe A
Instância 3: classes A and B
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Problema com única classe:

Classificador	Positiva	Negativa
A	1, 2, 3, 6	4,5
B	1, 3, 5	2, 4, 6
C	4	1, 2, 3, 5, 6

Classificação Multirrótulos (5)

- **Estratégia 2:** criar novos rótulos

Problema Multirrótulo:

Instância 1: classes A and B
Instância 2: classe A
Instância 3: classes A and B
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Problema monorrótulo:

Instância 1: classe D
Instância 2: classe A
Instância 3: classe D
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Classificação Multirrótulos ⁽⁶⁾

- **Estratégia 3:** eliminar instâncias

Problema Multi-rótulo:

Instância 1: classes A and B
Instância 2: classe A
Instância 3: classes A and B
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Problema Um-rótulo:

Instância 2: classe A
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Classificação Multirrótulos ⁽⁷⁾

- **Estratégia 4:** eliminar rótulos

Problema Multirrótulo:

Instância 1: classes A and B
Instância 2: classe A
Instância 3: classes A and B
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Problema Um-rótulo:

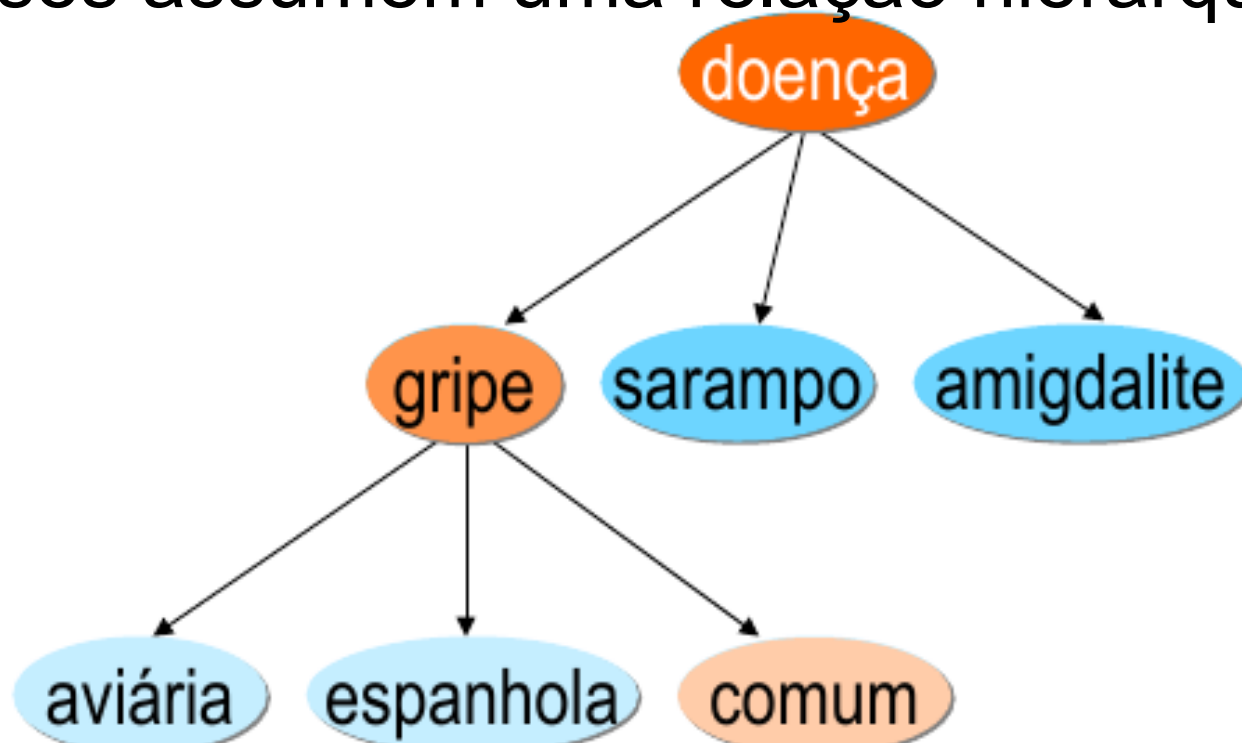
Instância 1: classe A
Instância 2: classe A
Instância 3: classe B
Instância 4: classe C
Instância 5: classe B
Instância 6: classe A

Classificação Hierárquica

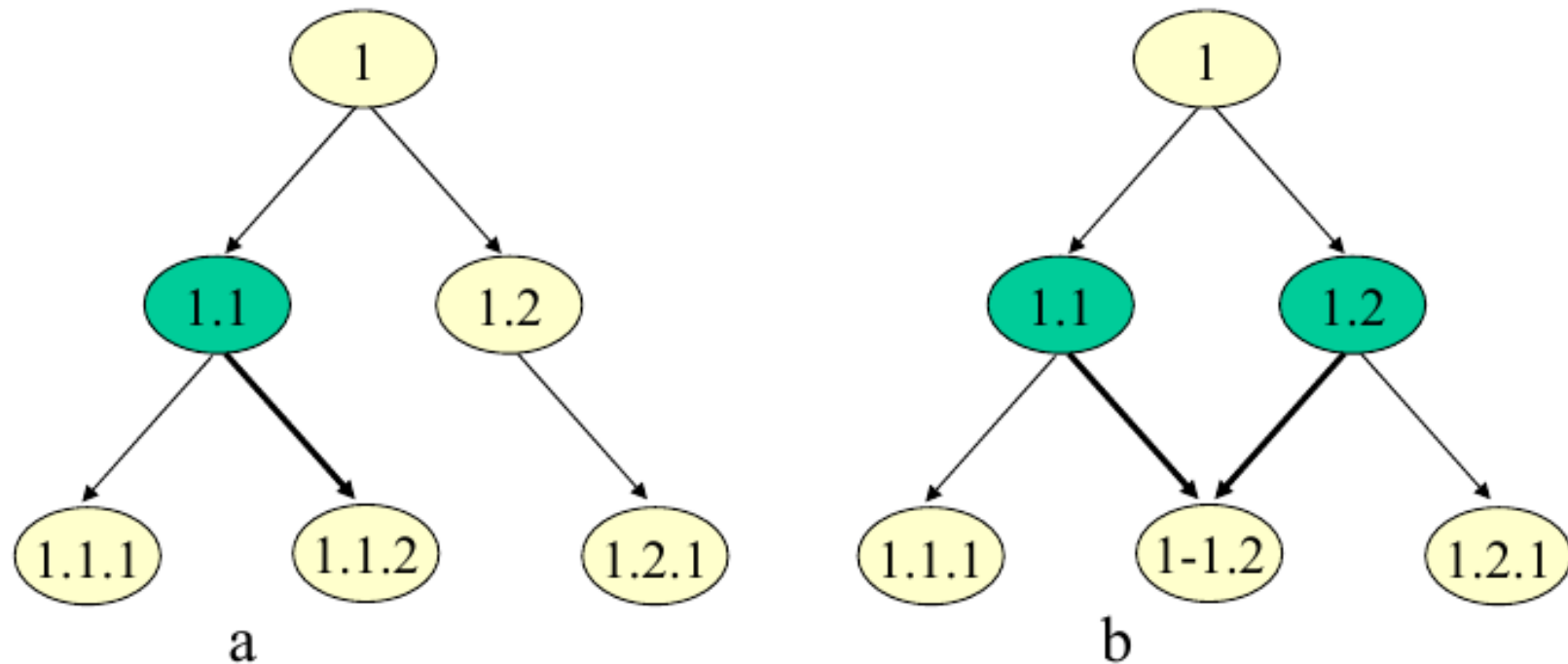
- Problemas de classificação em que:
 - Classes podem ser divididas em subclasses
 - Classes podem ser agrupadas em superclasse
- Classificação baseada em uma hierarquia
 - Predição obrigatória ou opcional em nós-folha

Classificação Hierárquica ⁽²⁾

- Dados hierarquicamente organizados
 - $\{1, 1.1, 1.2, \dots, k, k.1, k.2\}$
 - Classes assumem uma relação hierárquica



Tipos de Hierarquia



(a) Árvore

(b) Grafo Direcionado Acíclico (DAG)

Classificação Hierárquica

- Principais abordagens
 - Transformação em um problema de classificação plana (escolher um nível da árvore)
 - Um classificador por nível sem respeitar hierarquia
 - Top-down: um classificador por nível da raiz até as folhas, respeitando hierarquia
 - Big-bang: um classificador só que analisa a hierarquia como um todo

Classificação Hierárquica ⁽⁶⁾

- Avaliação
 - Custo uniforme
 - Mais utilizada
 - Custo baseado em distância
 - Baseada na distância entre classe predita e classe verdadeira
 - Custo baseado na semântica
 - Quanto mais similares as classes, menor a penalização

Classificação com Ranking

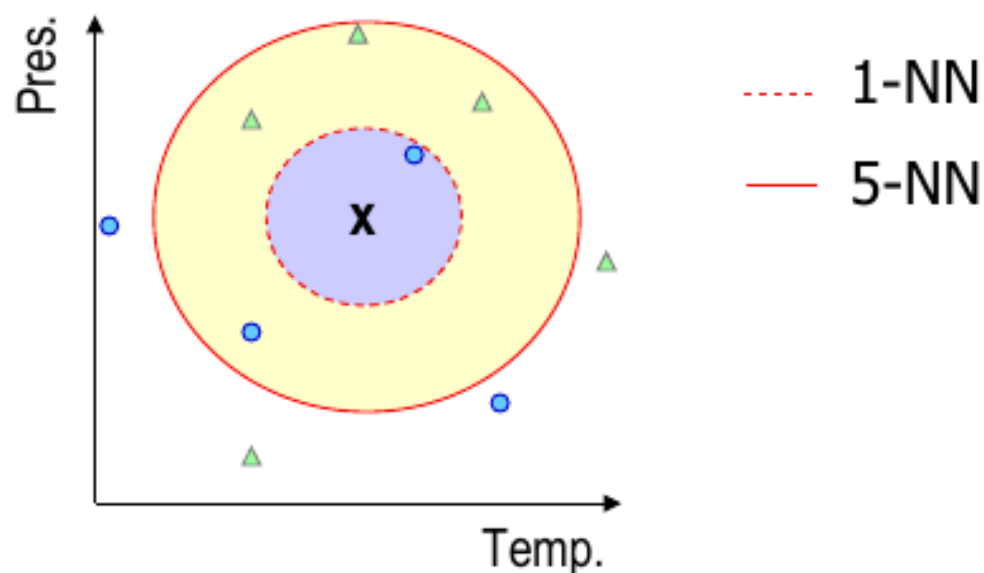
- Caso especial de classificação multirrótulos
 - Responde com as N classes que podem ser atribuídas ao exemplo de entrada
 - Ordenadas por relevância
 - Muito utilizada em recuperação de informação
 - Taxa de erro pode ser medida pela comparação de rankings

Técnicas de classificação

- Algumas técnicas de classificação:
 - k-vizinhos mais próximos
 - Regras de Decisão
 - Árvores de Decisão
 - Redes Neurais Artificiais
 - Support Vector Machines

k-vizinhos mais próximos

- Aprendizado baseado em instância
 - Classifica de acordo com distância aos vizinhos



Simple

Armazenamento de dados (não há modelo explícito)

k-vizinhos mais próximos ₍₂₎

Para cada novo exemplo
Definir a classe dos k exemplos mais próximos
Classificar exemplo na classe majoritária de seus vizinhos

k-vizinhos mais próximos ⁽³⁾

- Quantos vizinhos
 - K muito grande
 - Vizinhos podem ser muito diferentes
 - Predição tendenciosa para classe majoritária
 - Custo computacional mais elevado
 - K muito pequeno
 - Não usa informação suficiente
 - Previsão pode ser instável
- Distâncias podem ser ponderadas

k-vizinhos mais próximos ₍₄₎

- Testar no Weka: KNN com diferentes valores de k e diferentes distâncias

Regras de Decisão

- Organiza informações em regras do tipo:

se X então Y
senão Z

se temperatura > 37°C **e** pressão > 12.7
então Doente
senão Saudável

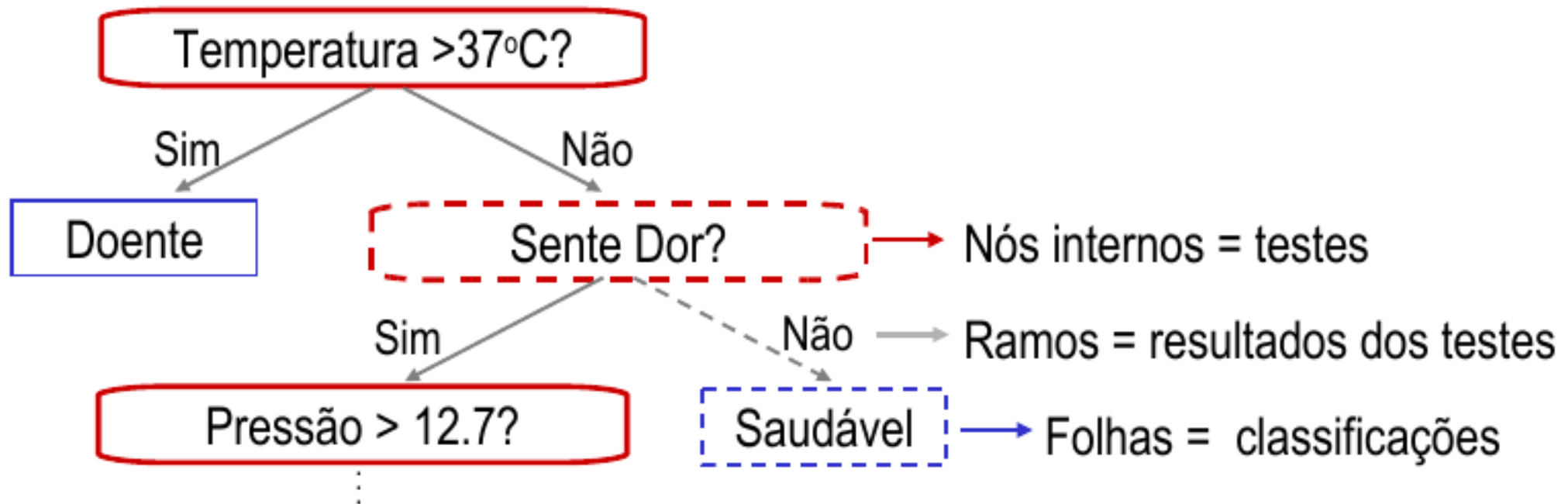
Simple

Dificuldade em lidar com dados contínuos

Pouca robustez a dados de grande dimensão

Árvore de Decisão

- Estrutura composta de nós e ramificações



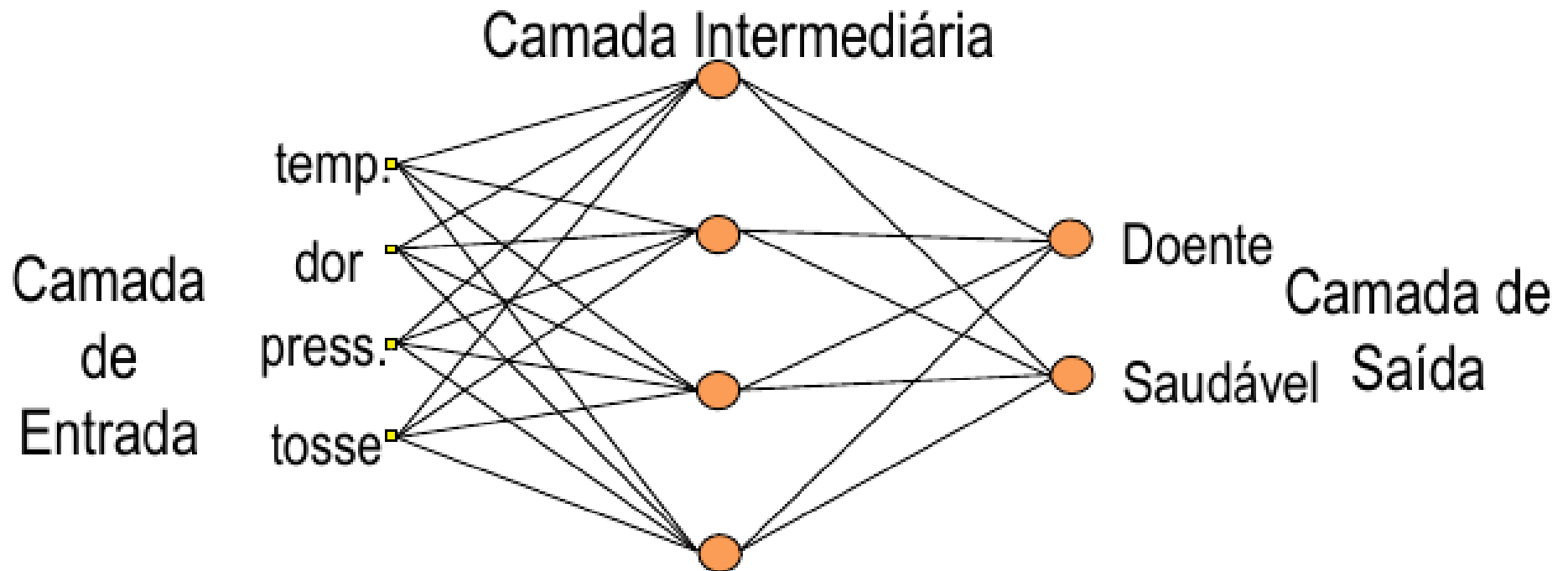
Compreensibilidade do modelo

Dificuldade em lidar com dados contínuos

Pouca robustez a dados de grande dimensão

Redes Neurais Artificiais

- Inspiradas na estrutura/funcionamento do cérebro

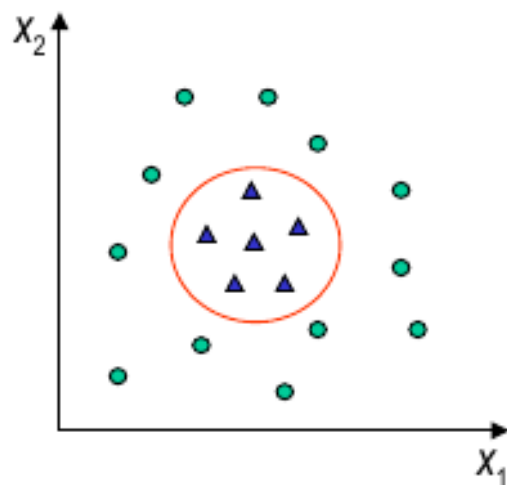


Capacidade de representar funções de formas variadas

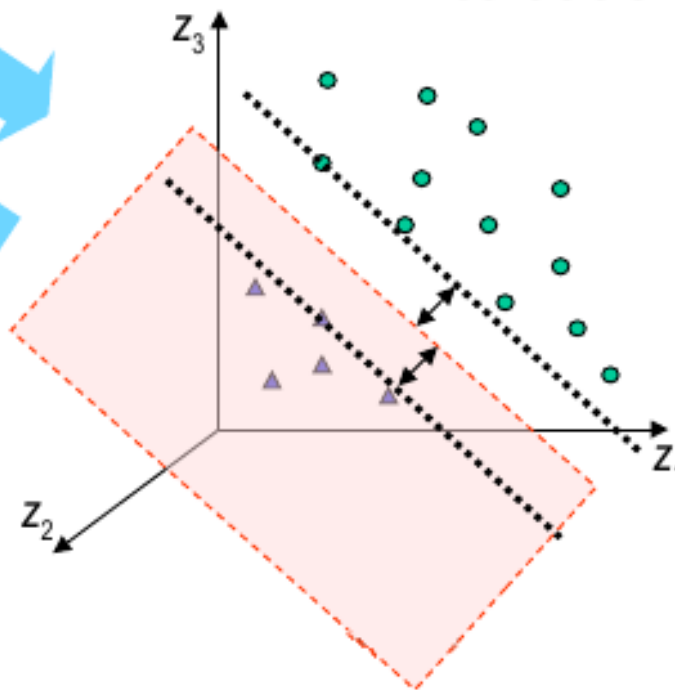
Ajuste de parâmetros

Dificuldade de interpretação do modelo

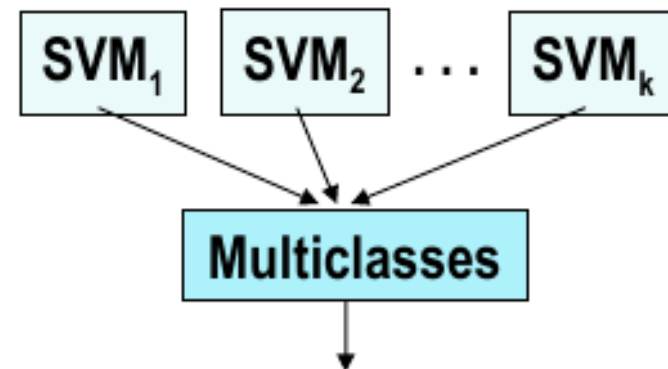
Support Vector Machines



Boa capacidade de generalização
Robustez a grandes dimensões
Ajuste de parâmetros
Dificuldade de interpretação do modelo




Classificação binária



Técnicas de Classificação

- Componentes algoritmos de classificação em AM:
 - Representação: viés (bias); super- e sub-ajustamento (over- e under-fitting)
 - Critério de avaliação: taxa de acerto; desempenho computacional; compreensibilidade
 - Método de busca: busca de parâmetros; busca do modelo



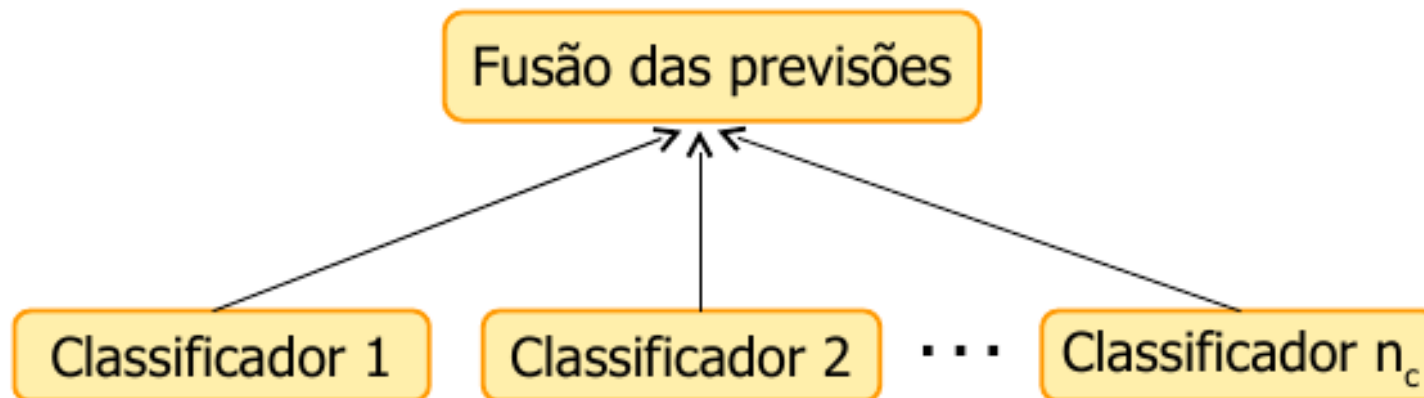
Não há técnica universal

Combinação de Classificadores

- **Binário** → **multiclasse** (homogêneo):
 - AAA, OAA, ECOC
- **Classificadores instáveis** (homogêneo):
Bagging, boosting
- **Classificadores diferentes** para obter recortes mais poderosos (heterogêneo)
 - Voto, média, outro classificador

Combinação de Classificadores ⁽²⁾

- Melhorar o desempenho
- Bagging
- Boosting



Combinação de Classificadores ⁽³⁾

- Bagging (Bootstrap Agregating):
 - Cada classificador é treinado com uma diferente amostra do conjunto de treinamento
 - Mesmo tamanho do conjunto original
 - Classe definida por votação
 - Filosofia: só existe um jeito de acertar; mas existem vários jeitos de errar
 - Na média, acerto é mais “votado” do que os vários erros

Combinação de Classificadores (4)

- Bagging :
 - Indicado para classificadores instáveis
 - Pequena mudança nos dados de treinamento afeta muito a classificação
 - Testar:
 - **weka.classifier.meta.bagging**

Combinação de Classificadores ₍₅₎

- Boosting
 - Conjunto de técnicas: ex. adaboost
 - Melhora desempenho de algoritmos fracos
 - Pouco melhores que classificação aleatória
 - Desempenho dos classificadores é influenciado pela escolha do conjunto de treinamento
 - Exemplos mais difíceis são escolhidos com maior probabilidade

Combinação de Classificadores (6)

- Boosting
 - Instável
 - Pouco indicado para dados com ruídos e pequenos conjuntos de dados
 - **`weka.classifier.meta.adaboost`**

Combinação de Classificadores ⁽⁷⁾

- Heterogêneo
 - Classificadores de diferentes abordagens
 - Treinados com
 - Todo conjunto de dados
 - Parte do conjunto de dados
- Avaliação
 - Voto: **weka.classifier.meta.vote**
 - Stacking: **weka.classifier.meta.stacking**
 - Outro classificador

Exercícios vistos em aula

- Rodar no Weka
- 1. Testar algoritmos de regressão para base Iris
- 2. Knn com diferentes valores de k e diferentes distâncias

Exercícios vistos em aula ₍₂₎

- 3. Meta-classificador multi-classe para classificadores baseados em duas classes (`weka.classifiers.meta.multiclassifier`)
 - Testar: AAA, OAO, ECOC
- 4. Meta-classificadores baseados em votação (`weka.classifiers.meta.vote`)

Pontos chaves

- Tipos de classificação (binária, uma classe, multiclasse, multi-rótulo, hierárquica, etc)
- K-vizinhos mais próximos (KNN)
- Espaços: de instância, de hipóteses e de versões
- Matrizes de código e combinação de classificadores
- Dimensão VC

Agradecimentos/referências

- Notas de aula do Prof. André de Carvalho (USP)