

Inteligência Artificial

Aprendizado supervisionado 1

Motivação

- Escrever um programa de computador que:
 - **Reconheça** pessoas pelo rosto
 - **Problemas:**
 - Diferentes expressões faciais
 - Alterações na face (ex. óculos, bigode)
 - Cortes de cabelo
 - Etc.



Que características considerar??

Seres humanos: reconhecimento de padrões,
aprendizado do que deve ser observado após vários exemplos

Motivação

- Escrever um programa de computador que:
 - Faça **diagnóstico** de pacientes por sintomas
 - **Médico**: formação e experiência
 - Responda a **questões** sobre vendas como:
 - Quais produtos são vendidos em **conjunto**?
 - Que produto **recomendar** a um cliente?
 - Como agrupar clientes para melhor **marketing**?



AM: outras motivações

- Automatizar o processo de aquisição de conhecimento
- Entender melhor os mecanismos de aprendizado humano
- Algumas tarefas são melhor definidas e/ou executadas a partir de exemplos
 - Ex.: Reconhecer pessoas
- Ser humano não é capaz de explicar (e programar) sua habilidade para executar alguns tipos de tarefas
 - Ex.: Dirigir



AM: outras motivações

- Quantidade de conhecimento disponível pode ser muito grande para ser descrito (e programado) por humanos
 - Ex.: diagnóstico médico
- Algumas tarefas exigem cálculos complexos, possíveis apenas com computador
 - Ex.: interrelacionar/correlacionar grandes quantidades de dados
- Modelos podem se adaptar a novas situações



Inteligência Artificial e AM

- Necessidade de ferramentas mais **autônomas**
 - Reduzindo necessidade de intervenção humana e dependência de especialistas

Aprendizado de Máquina: técnicas capazes de criar, a partir de experiência passada, uma hipótese (função) capaz de resolver o problema

- Ex: **Se** temperatura $> 37^{\circ}\text{C}$ **e** tem dores **então** está doente
 - Regra definida a partir de prontuários médicos



Aprendizado de Máquina: definição

Um programa **aprende** a partir da experiência **E**, em relação a uma classe de tarefas **T**, com medida de desempenho **P**, se seu desempenho em **T**, medido por **P**, melhora com **E**

Mitchell, 1997

Algoritmos de AM: induzem uma função ou hipótese capaz de resolver o problema a partir de instâncias do problema a ser resolvido

Exemplo 2

- **Problema:** filtrar mensagens de email
 - **Tarefa T:** categorizar mensagens de *email* como *spam* ou legítima
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



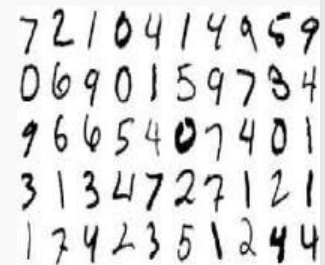
Exemplo 2

- **Problema:** filtrar mensagens de email
 - **Tarefa T:** categorizar mensagens de *email* como *spam* ou legítima
 - **Medida de desempenho P:** porcentagem de mensagens de *spam* corretamente identificadas
 - **Experiência E:** conjunto de e-mails apontados pelo usuário como *spams*



Exemplo 3

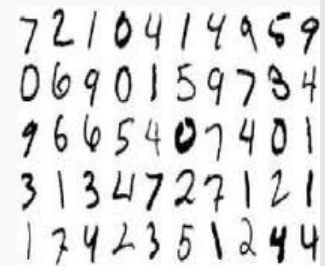
- **Problema:** reconhecer escrita manual
 - **Tarefa T:** reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



7210414959
0690159784
9665407401
3134727121
1742351244

Exemplo 3

- **Problema:** reconhecer escrita manual
 - **Tarefa T:** reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P:** porcentagem de dígitos corretamente identificados
 - **Experiência E:** exemplos de dígitos manuscritos por diferentes pessoas



7210414959
0690159784
9665407401
3134727121
1742351244

Exemplo 5

- **Problema:** diagnóstico médico
 - **Tarefa T:** diagnosticar o estado de um paciente dado um conjunto de sintomas
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



Exemplo 5

- **Problema:** diagnóstico médico
 - **Tarefa T:** diagnosticar o estado de um paciente dado um conjunto de sintomas
 - **Medida de desempenho P:** porcentagem de pacientes corretamente diagnosticados
 - **Experiência E:** prontuários médicos de pacientes



Exercício

Uma empresa quer fazer um sistema computacional que indique a que clientes uma promoção deve ser mandada, de maneira a ter uma melhor aderência (gerar mais compras)

Ela possui uma base de dados histórica de vendas anteriores do produto que será anunciado

Exercício: defina esse problema para o uso de Aprendizado de Máquina, dizendo qual é a **tarefa** (T), medida de **desempenho** (P) que pode ser utilizada no aprendizado e qual **experiência** (E) pode ser usada

Inferência Indutiva

- A **Inferência Indutiva** é um dos principais meios para a aquisição de novos conhecimentos
- **Indução**: raciocínio para obter conclusões sobre todos os membros de uma classe pelo exame de alguns membros da classe



Raciocínio do particular para o geral

Inferência indutiva: exemplo

- Se eu noto que:
 - Todos pacientes com déficit de atenção atendidos em 1986 sofriam de ansiedade
 - 95% pacientes com déficit de atenção atendidos em 1987 sofriam de ansiedade
 - ...
 - **⇒ Posso inferir que pacientes que sofrem de déficit de atenção também sofrem de ansiedade**

Isto pode ser ou não verdade,
mas propicia uma boa generalização

Conjunto de dados

- Experiência pode ser provida por um conjunto de dados (de treinamento)
 - Ex. base de dados de um hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Conjunto de dados

- Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Meta: induzir hipótese para fazer diagnósticos corretos para novos pacientes

Conjunto de dados

▪ Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Cada **linha** (paciente) é um **dado**
(objeto, exemplo, padrão ou registro)

Conjunto de dados

▪ Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Cada objeto é uma tupla com valores de **características** (atributos, campos ou variáveis), que descrevem seus principais aspectos

Conjunto de dados

▪ Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Atributo de saída (alvo/meta): presente em algumas tarefas, seus valores devem ser estimados usando outros atributos (de entrada/preditivos)

Importante: atributos de identificação e nome não possuem relação com a doença e não são utilizados como entradas

Conjunto de dados

■ Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	27	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	87	Uniformes	38,4	2	GO	Saudável

Importante: lidar com dados imperfeitos (ruídos, ausentes, etc.)

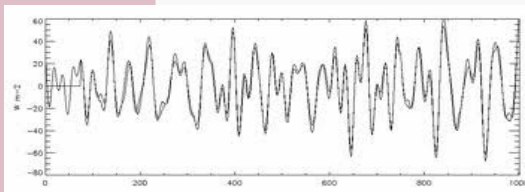
Dados

- Estima-se que a quantidade de dados em Bases de Dados mundiais dobra a cada 20 anos
- Crescimento tem ocorrido em várias áreas
 - Transações bancárias
 - Utilização de cartões de crédito
 - Dados governamentais
 - Medições ambientais
 - Dados clínicos
 - Projetos genoma
 - Informações disponíveis na web
 - etc.



Dados

- Podem ter diferentes formatos



Séries temporais

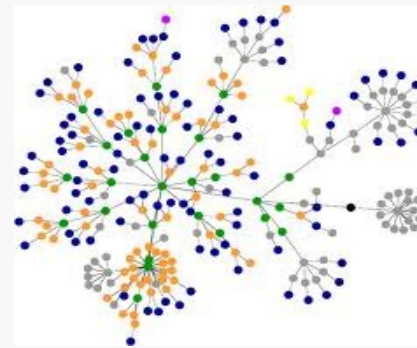


Páginas web

Die noch lebhafteste Partenonstrate sichtlich auffallende Art ist besonders durch die Neigung der kalten Bienenmischung zur Adhäsion in eine unvollständige Endform, sowie das Auftreten einer kalten Intervall- und verhältnis zugewandten inneren Endformel mehr ausgezeichnet. Ähnlich wie bei den Bienen D. Gressitt tritt, steht sich eine nur stärker erhaltene Längsrippe, die etwas innerhalb der Schärfe beginnt, bis über die Mitte. Im Gegensatz zu dieser Art ist Kopf und Hinterkopf in viel geringer Ausdehnung nicht entfernt, da die schmalen Mittelrinne und besonders die viel weniger ausgeprägte Seitenrinne mehr Raum für die Bildung übrig lassen. Diese ist sehr klein, nur den Hinterkopf zwischen den kalten Seiten auf der Seite leicht oberhalb oder unterhalb, der dritte Teil des Kopfes, wenn eine einzelne Keilung der glatten/geraden Mittelrinne des Thorax und der Rumpfes ausstrahlt der Seitenrinne verläuft. Das Geschehen der Flügeladern ist heller als dunkler lauffaden, sie selbst abgrenzen weitere Ränder der ÖÖ tritt unvollständig begrenzt, teils in eine Reihe von Flecken auftritt, während der Gesamtstruktur von den der Ringe spärlichen Divisionen mit ihnen selbst begrenzt, teils durch wesentlich absteigt. Von D. Thagot Press und Medinat Press unterteilt sich aufpassen, abgesehen von der Zeichnung, durch die bei diesen beiden Arten ganz ähnlichen oder umgekehrten Schwebelängen des Halses. Bei D. Gressitt tritt, ist die kalte Halskette tiefer gefaltet und die Kantenlinie der Flügeladern sehr schnell, sondern fehlt bei dieser Art die Rückengeige.

Siquen *) in Casteln (Karl, 17. 8. 87)

Textos



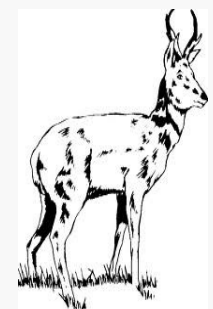
Grafos



Áudios



Vídeos



Imagens

Geralmente transformados para o formato atributo-valor

Conjunto de dados

- Pode ser representado por uma matriz de objetos $\mathbf{X}_{n \times d}$
 - n = número de objetos
 - d = número de atributos (excluindo atributo-meta)
 - Dimensionalidade dos objetos
 - Do [espaço de objetos](#) (de entradas/de atributos)
 - Elemento x_i^j (ou x_{ij}) \Rightarrow valor da j -ésima característica para o objeto i

Conjunto de dados: exercício

Considere que temos o seguinte conjunto de dados da empresa de marketing:

<i>Id_cliente</i>	<i>Rendimento</i>	<i>Idade</i>	<i>Tamanho da família</i>	<i>Já comprou outros produtos da empresa</i>	<i>Comprou o produto anunciado</i>
C1	Baixo	> 40	Grande	Sim	Não
C2	Baixo	> 40	Grande	Não	Não
C3	Alto	> 40	Grande	Sim	Sim
C4	Médio	[30,40]	Grande	Sim	Sim
C5	Médio	< 30	Pequena	Sim	Sim
C6	Médio	< 30	Pequena	Não	Não
C7	Alto	< 30	Pequena	Não	Sim
C8	Baixo	[30,40]	Grande	Sim	Não
C9	Baixo	< 30	Pequena	Sim	Sim
C10	Médio	[30,40]	Pequena	Sim	Sim
C11	Baixo	[30,40]	Pequena	Não	Sim
C12	Alto	[30,40]	Grande	Não	Sim
C13	Alto	> 40	Pequena	Sim	Sim
C14	Médio	[30,40]	Grande	Não	Não

Exercício:

Responda:

O que representam os objetos nesse conjunto?

E os atributos?

Todos são úteis para o aprendizado?

Tipos de atributos

Quantitativo (numérico)

Representa quantidades

Valores podem ser ordenados
e usados em operações
aritméticas

Podem ser **contínuos ou**
discretos

Possuem unidade associada

Qualitativo (simbólico ou categórico)

Representa qualidades

Valores podem ser associados
a categorias

Alguns podem ser ordenados,
mas operações aritméticas
não são aplicáveis

Ex. {pequeno, médio, grande}

Tipos de atributos

Atributos Quantitativos

Contínuos

- Podem assumir um número **infinito** de valores
- Geralmente resultados de medidas
- Frequentemente representados por **números reais**
- *Ex. peso, distância*

Discretos

- Número **finito ou infinito contável** de valores
- Caso especial: atributos binários (booleanos)
- *Ex. {12, 23, 45}, {0, 1}*

Tipos de atributos

- Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Tipos de atributos

- Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Qualitativo
contínuo

Quantitativo discreto

Quantitativo

Tipos de atributos

- Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Alguns atributos qualitativos são representados por números, mas não faz sentido a utilização de operadores aritméticos sobre seus valores

Escala de atributos

- Define operações que podem ser realizadas sobre os valores dos atributos
 - Nominais
 - Ordinais
 - Intervalares
 - Racionais

Escala de atributos

- Define operações que podem ser realizadas sobre os valores dos atributos
 - Nominiais
 - Ordinais
 - Intervalares
 - Racionais

Qualitativos

Escala de atributos

- Define operações que podem ser realizadas sobre os valores dos atributos
 - Nominiais
 - Ordinais
 - Intervalares
 - Racionais

Quantitativos

Escalas de atributos

Escala nominal

- Valores são nomes diferentes e carregam a menor quantidade de informação possível
- Não existe relação de ordem entre os valores
- **Operações aplicáveis:** =, ≠
- *Ex.: número de conta em banco, cores, sexo*

Escala ordinal

- Valores refletem ordem das categorias representadas
- **Operações aplicáveis:** =, ≠, <, >, ≤, ≥
- *Ex.: hierarquia militar, avaliações qualitativas de temperatura*



Escalas de atributos

Escala intervalar

- Números que variam em um intervalo
- É possível definir ordem e **diferença em magnitude** entre dois valores
- Origem da escala definida de maneira arbitrária
- **Operações aplicáveis:** =, ≠, <, >, ≤, ≥, +, -
- *Ex.: temperatura em °C ou °F, datas*

Escala racional

- Carregam mais informações
- Têm significado absoluto (**existe 0 absoluto**)
- Razão tem significado
- **Operações aplicáveis:** =, ≠, <, >, ≤, ≥, +, -, *, /
- *Ex.: tamanho, distância, salário, saldo em conta*



Escalas de atributos

- Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Escalas de atributos

- Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Nominal

Ordinal

Intervalar

Racional

Exercício

- Definir o tipo e escala dos seguintes atributos:
 - Renda mensal: ?
 - Número de palavras de um texto: ?
 - Número de matrícula: ?
 - Data de nascimento: ?
 - Código postal: ?
 - Posição em uma corrida: ?

Exercício

- Definir o tipo e escala dos seguintes atributos:
 - Renda mensal: **quantitativo racional**
 - Número de palavras de um texto: **quantitativo racional**
 - Número de matrícula: **qualitativo nominal**
 - Data de nascimento: **quantitativo intervalar**
 - Código postal: **qualitativo nominal**
 - Posição em uma corrida: **qualitativo ordinal**

Tarefas de Aprendizado

- Preditivas vs Descritivas

Previsão

Encontrar função (modelo ou hipótese) que possa ser utilizada para prever um rótulo ou valor para novos dados

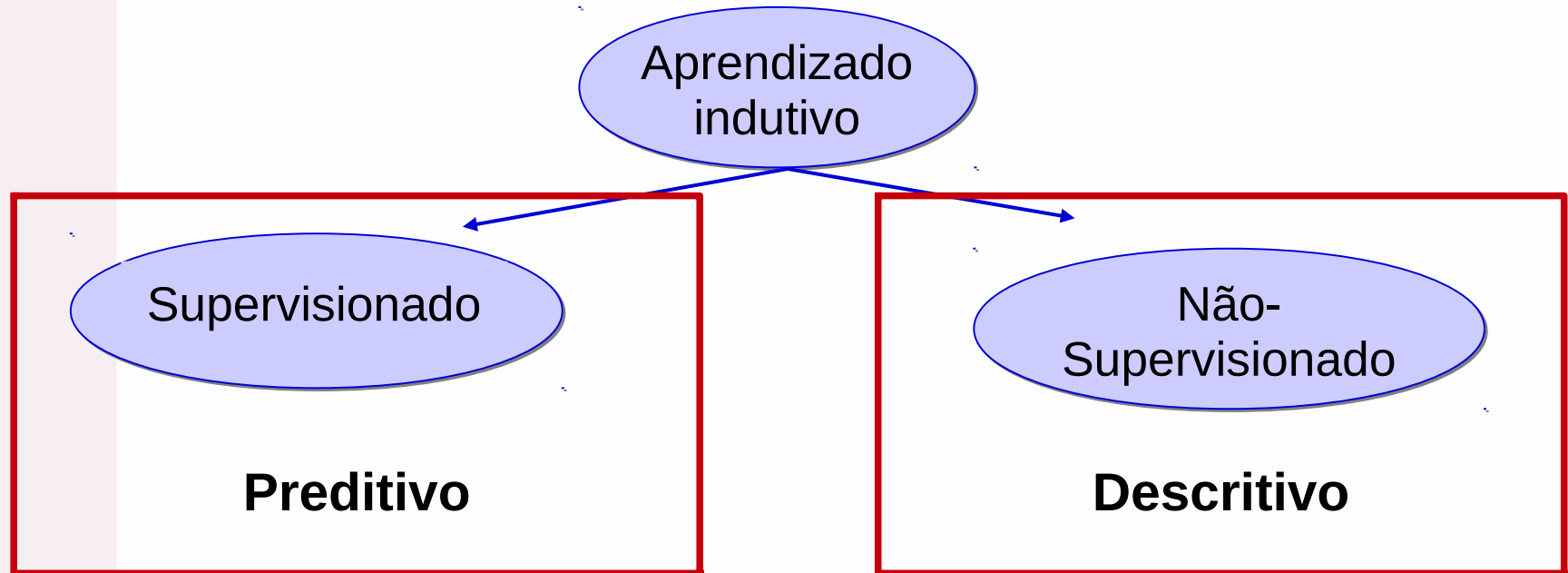
Objetos na forma
(entrada, saída)

Descrição

Explorar ou descrever um conjunto de dados

Objetos não possuem saída associada

Hierarquia de aprendizado



Importante: divisão não é rígida (modelo preditivo também provê descrição dos dados e modelo descritivo pode prover previsões após validado)

Aprendizado supervisionado

- Supervisor externo
 - Conhece saída desejada para cada exemplo
 - Representado por conjunto de pares (\mathbf{x}, y)
 - Ex.: \mathbf{x} = sintomas e y = diagnóstico

Classificação

Rótulos discretos

Ex.: diagnóstico, bom/mau pagador, etc.

Regressão

Rótulos contínuos

Ex.: peso, altura, etc.

Aprendizado não supervisionado

- Algoritmos não fazem uso de atributo de saída
 - Exploram regularidades nos dados

Sumarização

Encontrar descrição compacta para dados

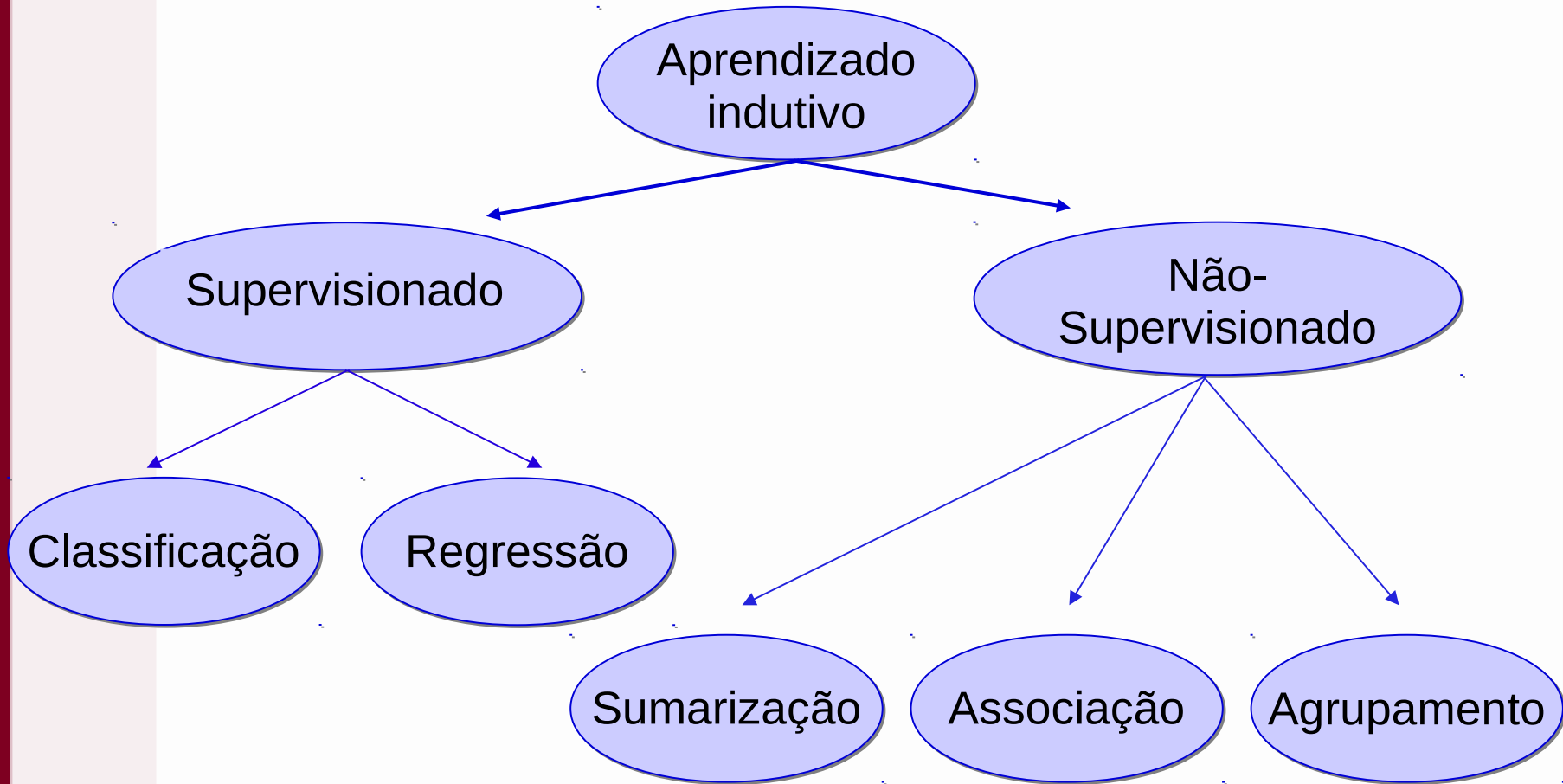
Associação

Encontrar padrões frequentes de associações entre atributos

Agrupamento

Dados agrupados de acordo com sua similaridade

Hierarquia de aprendizado



Aprendizado por reforço

- Reforçar/recompensar ações positivas e punir ações negativas
 - Crítico externo

Aprendizado supervisionado

- Supervisor
- É dito o que fazer
- Mais rápido

Aprendizado por reforço

- Crítico
- Faz e vê o que acontece
- Mais lento

Pacotes e conjuntos de dados

- UCI Machine Learning repository
 - <http://archive.ics.uci.edu/ml/>
- Weka
 - <http://www.cs.waikato.ac.nz/ml/weka/>
- Scikit-learn
 - <http://scikit-learn.org/stable/>
- R Project
 - <http://www.r-project.org/>

Aprendizado Supervisionado

Métodos Baseados em Distâncias

Modelos preditivos

- Algoritmo de AM preditivo: função que, dado um conjunto de exemplos rotulados, constrói um estimador

Classificação

- Rótulos nominais (conjunto discreto e não ordenado de valores)

Ex. {doente, saudável}, {bom pagador, mau pagador}, {iris setosa, iris versicolor, iris virginica}

- Estimador é chamado

classificador

Regressão

- Rótulos contínuos (conjunto infinito ordenado de valores)

Ex. peso, temperatura, vazão de água

- Estimador é chamado

regressor

Estimadores podem ser vistos como funções

Modelos preditivos

- Definição formal: dado conjunto de observações $\mathbf{D} = \{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$
 - Pode-se considerar que $y_i = f(\mathbf{x}_i)$, em que f representa uma função desconhecida
 - Mapeia entradas em saídas correspondentes
 - Algoritmo preditivo aprende **aproximação** f^*
 - Que permite estimar valor de y para novos objetos \mathbf{x}

Classificação

$$y_i = f^*(\mathbf{x}_i) \in \{c_1, \dots, c_m\}$$

Regressão

$$y_i = f^*(\mathbf{x}_i) \in \mathcal{R}$$

Modelos preditivos

- Exemplos de conjuntos de dados:

Conjunto de dados iris

TamP	LargP	TamS	LargS	Espécie
5,1	3,5	1,4	0,2	Setosa
4,9	3	1,4	0,2	Setosa
7	3,2	4,7	1,4	Versicolor
6,4	3,2	4,5	1,5	Versicolor
6,3	3,3	6	2,5	Virgínica
5,8	2,7	5,1	1,9	Virgínica

Classificação

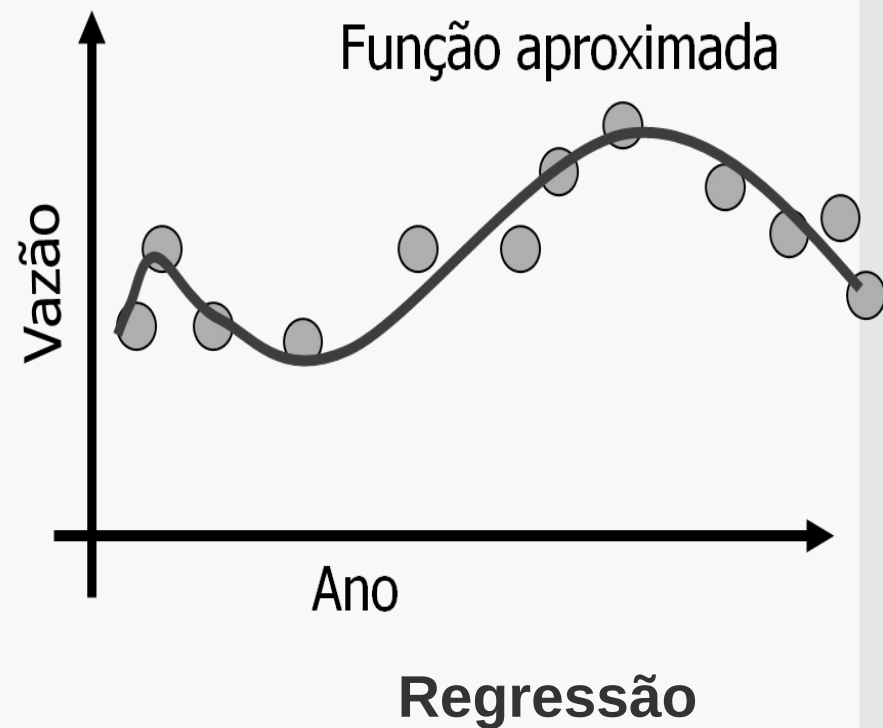
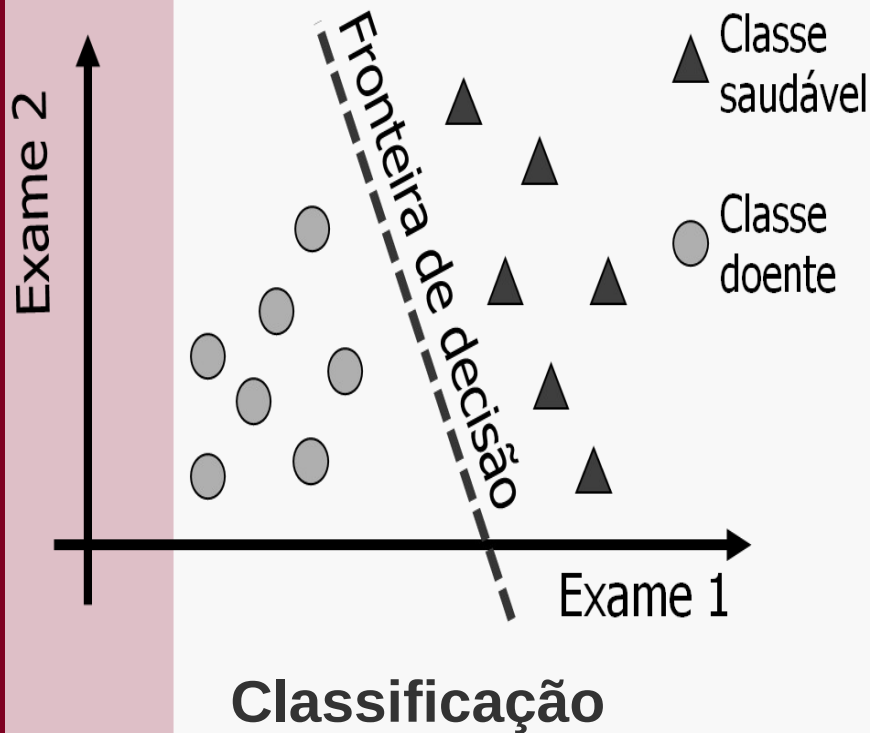
Conjunto de dados swiss

Fertilidade	Agricultura	Educação	Renda	Mortalidade
80,2	17	12	9,9	22,2
83,1	45,1	9	84,8	22,2
92,5	39,7	5	93,4	20,2
85,8	36,5	7	33,7	20,3
76,9	43,5	15	5,2	20,6

Regressão

Modelos preditivos

- Ilustração das tarefas:



Exercício

- Identifique se cada um dos seguintes problemas são de **classificação** ou **regressão**
 - Diagnosticar se um paciente tem uma doença ou não: ?
 - Prever o valor de mercado de um imóvel: ?
 - Dizer de que assunto uma notícia trata: ?
 - Dizer qual é o caractere manuscrito em uma imagem: ?
 - Prever tempo de desgaste de peça: ?
 - Fundir as leituras de vários sensores robóticos em um valor só: ?

Exercício

- Identifique se cada um dos seguintes problemas são de **classificação** ou **regressão**
 - Diagnosticar se um paciente tem uma doença ou não: **classificação**
 - Prever o valor de mercado de um imóvel: **regressão**
 - Dizer de que assunto uma notícia trata: **classificação**
 - Dizer qual é o caractere manuscrito em uma imagem: **classificação**
 - Prever tempo de desgaste de peça: **regressão**
 - Fundir as leituras de vários sensores robóticos em um valor só: **regressão**

Métodos baseados em distâncias

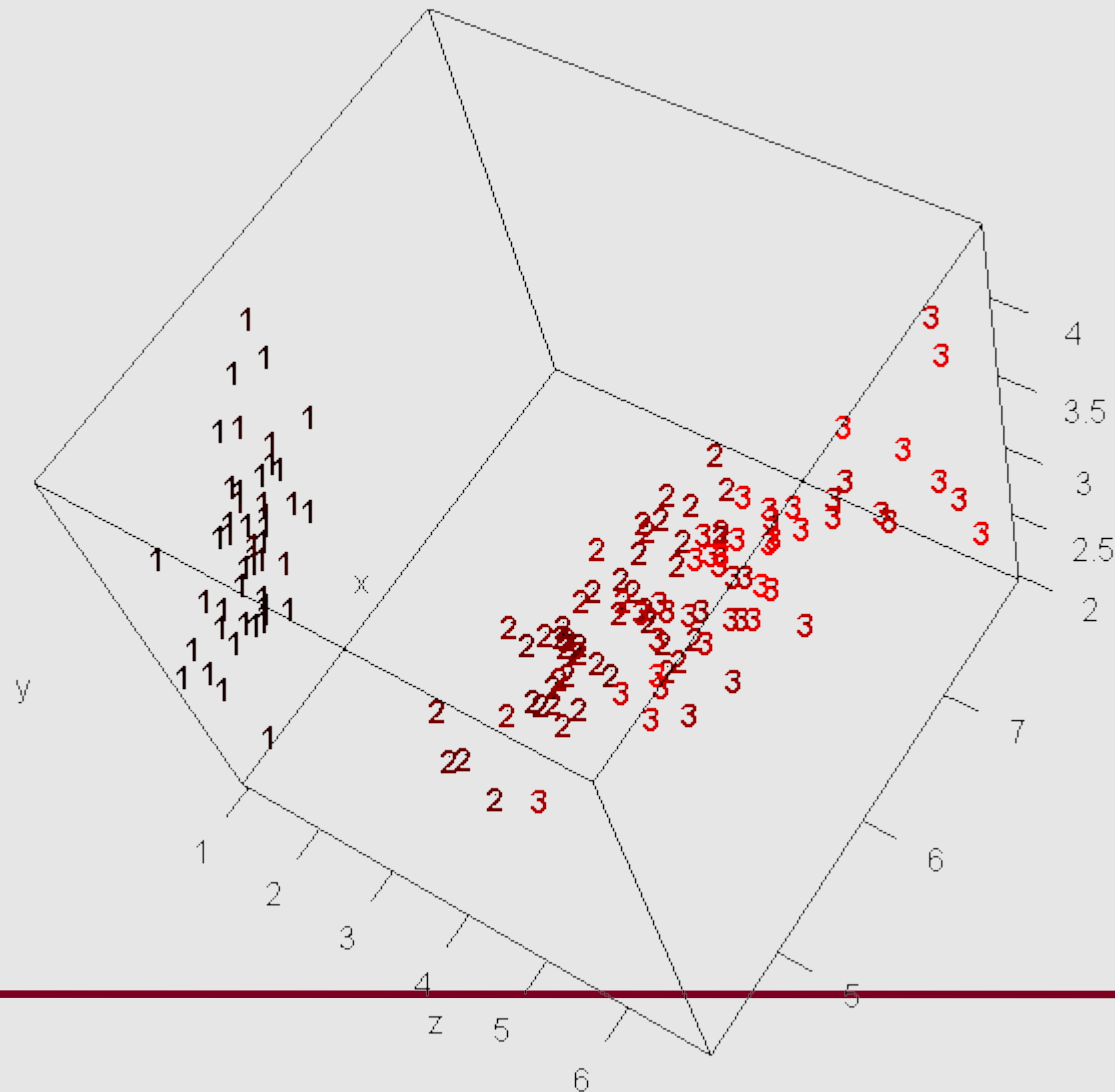
- Técnicas de AM que consideram proximidade entre os dados para realizar previsões

Hipótese: dados similares tendem a estar concentrados em uma mesma região do espaço de entradas

- E dados que não são similares estarão distantes entre si

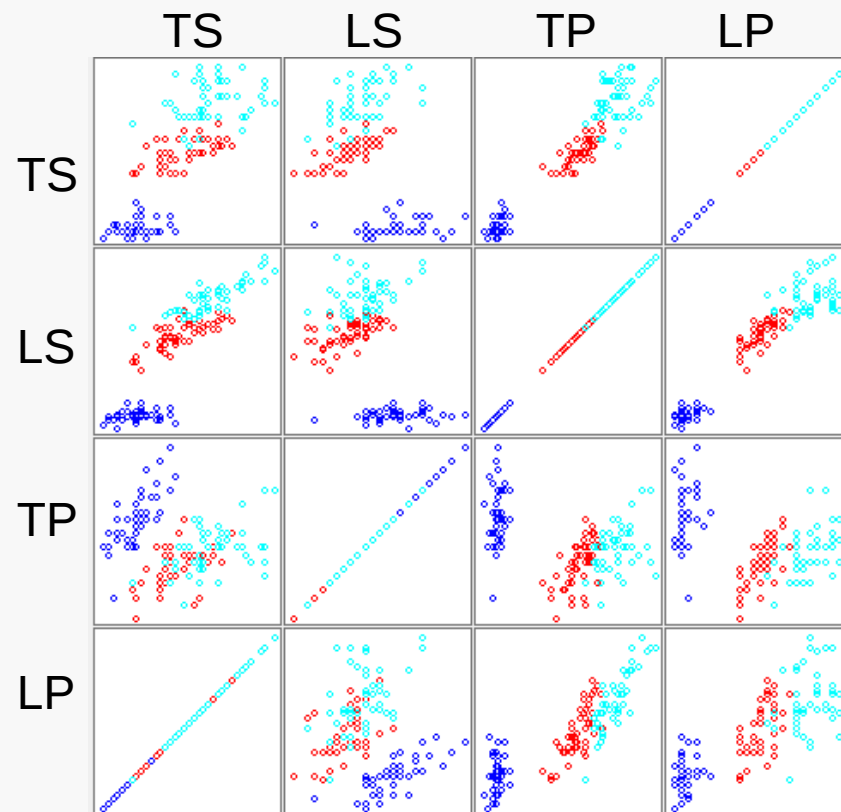
Métodos baseados em distâncias

- Ex. conjunto de dados iris
 - x = comp. da sépala
 - y = larg. da sépala
 - z = comp. da pétala
 - Cor = larg. da pétala
 - Números = classe



Métodos baseados em distâncias

- Ex. conjunto de dados iris



Proximidade

- Medida de **proximidade** entre pares de objetos pode ser de:

Similaridade

- Mede o quanto dois objetos são **parecidos**
- Quando **mais parecidos** \Rightarrow **maior o valor**
- Geralmente valor $\in [0, 1]$

Dissimilaridade

- Mede o quanto dois objetos são **diferentes**
- Quando **mais diferentes** \Rightarrow **maior o valor**
- Geralmente valor $\in [0, X]$

Escolha da medida deve considerar tipos e escalas dos atributos, além de propriedades dos dados que se deseja focar

Medidas para atributos quantitativos: dissimilaridade

- Para atributos quantitativos, mais usadas são distâncias baseadas na métrica de Minkowski
 - Distância L_p , $1 \leq p < \infty$

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt[p]{\sum_{l=1}^d |x_i^l - x_j^l|^p}$$

- São sensíveis a variações de escala dos atributos
Normalmente solucionado por normalização

Distância Mahattan

- Distância de Manhattan:
 - Minkowski com $p = 1$
 - Também chamada distância bloco-cidade

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^d |x_i^l - x_j^l|$$

Distância Euclidiana

- Distância Euclidiana:
 - Minkowski com $p = 2$
 - Medida de distância mais popular

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^d (x_i^l - x_j^l)^2}$$

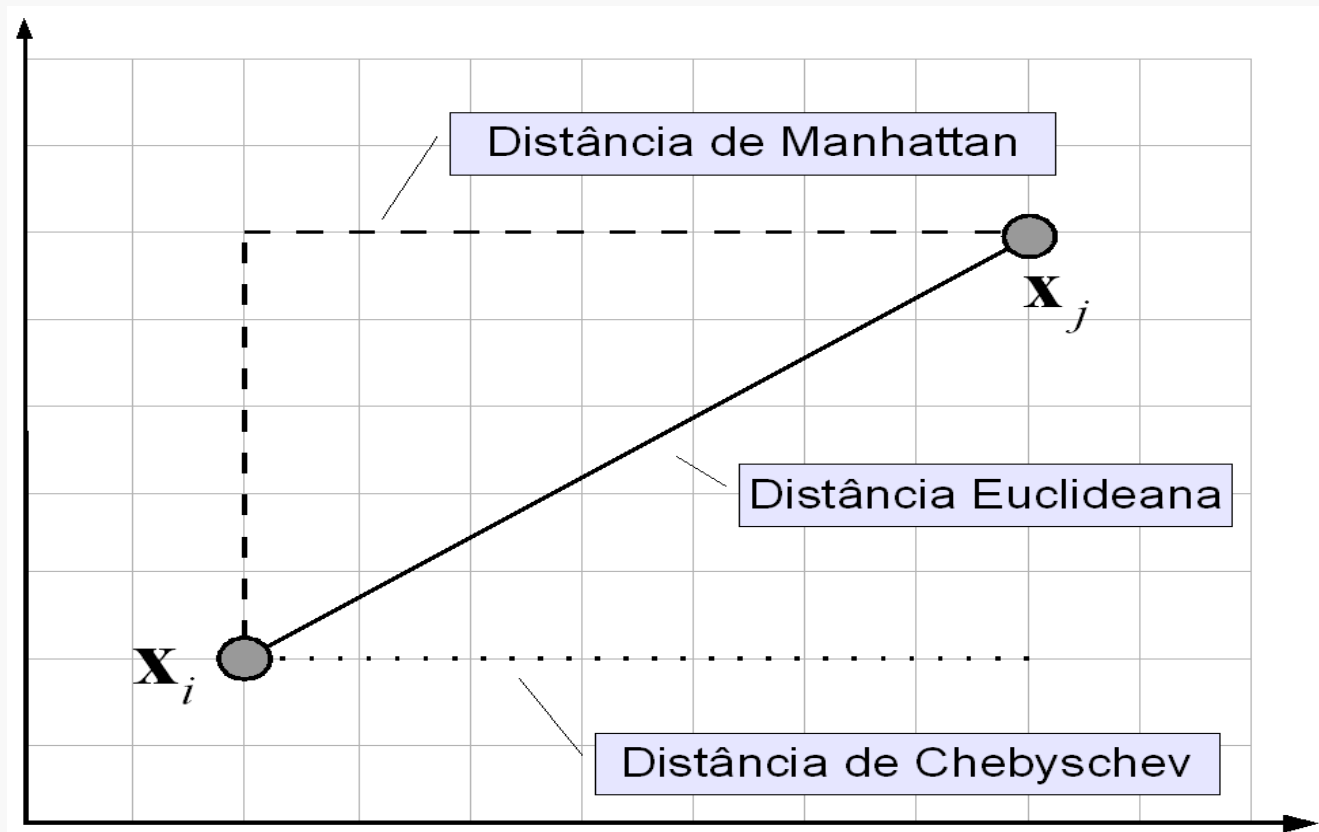
Distância Supremum

- Distância Supremum:
 - Minkowski com $p = \infty$
 - Também chamada distância de Chebyshev
 - Diferença absoluta máxima entre quaisquer atributos

$$d(\mathbf{x}_i, \mathbf{x}_j) = \max_l |x_i^l - x_j^l|$$

Métricas de Minkowski

- Interpretação das métricas de Minkowski para $d = 2$



Transformação de Dados

- Para usar as medidas de distância anteriores, todos os atributos devem ser quantitativos
 - Atributos qualitativos (simbólicos) devem ser convertidos
- Além disso, é necessário **normalizar** os atributos

Quando vários atributos estão em escalas diferentes

Para evitar que um atributo predomine sobre outro

Conversão simbólico-numérico

- Atributo simbólico com dois valores
 - Um dígito binário é suficiente
 - Ex. presença/ausência = 1/0
 - Se ordinal, 0 indica o menor valor e 1 o maior valor
- Atributo simbólico com mais valores
 - Conversão depende se o atributo é **nominal** ou **ordinal**

Conversão simbólico-numérico

- Ex. conjunto de dados hospital
 - Conversão de atributo Sexo para numérico

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	0	79	Grandes	38,0	2	Doente
18	1	67	Pequenas	39,5	4	Doente
49	0	92	Grandes	38,0	2	Saudável
18	0	43	Grandes	38,5	20	Doente
21	1	52	Médias	37,6	1	Saudável
22	1	72	Pequenas	38,0	3	Doente
19	1	87	Grandes	39,0	6	Doente
34	0	67	Médias	38,4	2	Saudável

M = 0

F = 1

Conversão simbólico-numérico

- Atributo nominal com mais valores
 - Inexistência de relação de ordem deve ser mantida
 - Diferença entre quaisquer dois valores numéricos deve ser a mesma
 - **Codificação canônica**: uso de c bits para c valores
 - Cada posição na sequência binária corresponde a um valor possível do atributo nominal
 - Cada sequência possui apenas um bit com valor 1
 - Distância de Hamming entre quaisquer dois valores é 2

Conversão simbólico-numérico

- Atributo nominal com mais que dois valores
 - Ex. codificação canônica (1-para-c ou topológica)

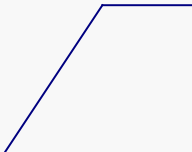
Atributo	Código 1-para-c
Azul	100000
Amarelo	010000
Verde	001000
Preto	000100
Marrom	000010
Branco	000001

Dependendo do número de valores nominais, pode gerar cadeias muito grandes de bits. Ex.: 193 nomes de países
Como resolver?

Conversão simbólico-numérico

- Atributo ordinal com mais que dois valores
 - Relação de ordem deve ser preservada
 - Ordenar valores ordinais e codificar cada um de acordo com sua posição na ordem com inteiro ou real

Atributo	Valor inteiro
Primeiro	0
Segundo	1
Terceiro	2
Quarto	3
Quinto	4
Sexto	5



Distância entre valores
varia de acordo com
proximidade entre eles

Conversão simbólico-numérico

- Ex. conjunto de dados hospital
 - Conversão de atributo ordinal Manchas

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	3	38,0	2	Doente
18	F	67	1	39,5	4	Doente
49	M	92	3	38,0	2	Saudável
18	M	43	3	38,5	20	Doente
21	F	52	2	37,6	1	Saudável
22	F	72	1	38,0	3	Doente
19	F	87	3	39,0	6	Doente
34	M	67	2	38,4	2	Saudável

Grandes = 3
Médias = 2
Pequenas = 1

Normalização por reescala

- Reescalar: adicionar/subtrair/multiplicar/dividir por uma constante
- Normalização min-max
 - São definidos inicialmente mínimo e máximo para os novos valores
 - Depois, para cada atributo aplica:

$$v_{\text{novo}} = \min + \frac{v_{\text{atual}} - \text{menor}}{\text{maior} - \text{menor}} \cdot (\text{max} - \min)$$

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de Idade entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Grandes	38,0	2	Doente
18	F	67	Pequenas	39,5	4	Doente
49	M	92	Grandes	38,0	2	Saudável
18	M	43	Grandes	38,5	20	Doente
21	F	52	Médias	37,6	1	Saudável
22	F	72	Pequenas	38,0	3	Doente
19	F	87	Grandes	39,0	6	Doente
34	M	67	Médias	38,4	2	Saudável

Maior = 49

Menor = 18

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de Idade entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Grandes	38,0	2	Doente
18	F	67	Pequenas	39,5	4	Doente
49	M	92	Grandes	38,0	2	Saudável
18	M	43	Grandes	38,5	20	Doente
21	F	52	Médias	37,6	1	Saudável
22	F	72	Pequenas	38,0	3	Doente
19	F	87	Grandes	39,0	6	Doente
34	M	67	Médias	38,4	2	Saudável

$$V_{\text{novo}} = \frac{V_{\text{atual}} - 18}{49 - 18}$$

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de Idade entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
0,32	M	79	Grandes	38,0	2	Doente
0	F	67	Pequenas	39,5	4	Doente
1	M	92	Grandes	38,0	2	Saudável
0	M	43	Grandes	38,5	20	Doente
0,1	F	52	Médias	37,6	1	Saudável
0,13	F	72	Pequenas	38,0	3	Doente
0,03	F	87	Grandes	39,0	6	Doente
0,52	M	67	Médias	38,4	2	Saudável

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de # Int. entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
0,32	M	79	Grandes	38,0	2	Doente
0	F	67	Pequenas	39,5	4	Doente
1	M	92	Grandes	38,0	2	Saudável
0	M	43	Grandes	38,5	20	Doente
0,1	F	52	Médias	37,6	1	Saudável
0,13	F	72	Pequenas	38,0	3	Doente
0,03	F	87	Grandes	39,0	6	Doente
0,52	M	67	Médias	38,4	2	Saudável

Maior = 20
Menor = 1

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de # Int. entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
0,32	M	79	Grandes	38,0	2	Doente
0	F	67	Pequenas	39,5	4	Doente
1	M	92	Grandes	38,0	2	Saudável
0	M	43	Grandes	38,5	20	Doente
0,1	F	52	Médias	37,6	1	Saudável
0,13	F	72	Pequenas	38,0	3	Doente
0,03	F	87	Grandes	39,0	6	Doente
0,52	M	67	Médias	38,4	2	Saudável

$$v_{\text{novo}} = \frac{v_{\text{atual}} - 1}{20 - 1}$$

Normalização por reescala

- Ex. conjunto de dados hospital
 - Normalização de # Int. entre 0 (min) e 1 (max)

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
0,32	M	79	Grandes	38,0	0,05	Doente
0	F	67	Pequenas	39,5	0,16	Doente
1	M	92	Grandes	38,0	0,05	Saudável
0	M	43	Grandes	38,5	1	Doente
0,1	F	52	Médias	37,6	0	Saudável
0,13	F	72	Pequenas	38,0	0,11	Doente
0,03	F	87	Grandes	39,0	0,26	Doente
0,52	M	67	Médias	38,4	0,05	Saudável

Normalização Z-score

Esta normalização permite, além de minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis, **centralizar os dados em zero**.

Valores entre -1 e 1.

$$\mu(x_j) := \frac{1}{m} \sum_{i=1}^m x_j^{(i)}$$

$$x_j^{(i)'} := \frac{x_j^{(i)} - \mu(x_j)}{\sigma(x_j)}$$

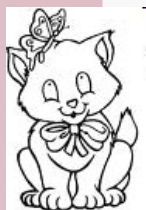
$$\sigma(x_j) := \sqrt{\frac{1}{m} \sum_{i=1}^m (x_j^{(i)} - \mu(x_j))^2}$$

Para cada ponto da característica ***j***

Algoritmo dos vizinhos mais próximos

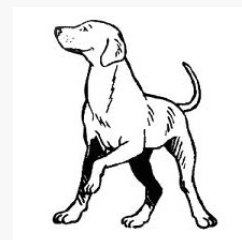
- Algoritmo de AM mais simples
 - **Intuição:** *Objetos relacionados ao mesmo conceito são semelhantes entre si*

Dados de
treinamento



distância

Exemplo
de teste



Se anda como um cachorro e late
como um cachorro, então
provavelmente é um ...

Algoritmo dos vizinhos mais próximos

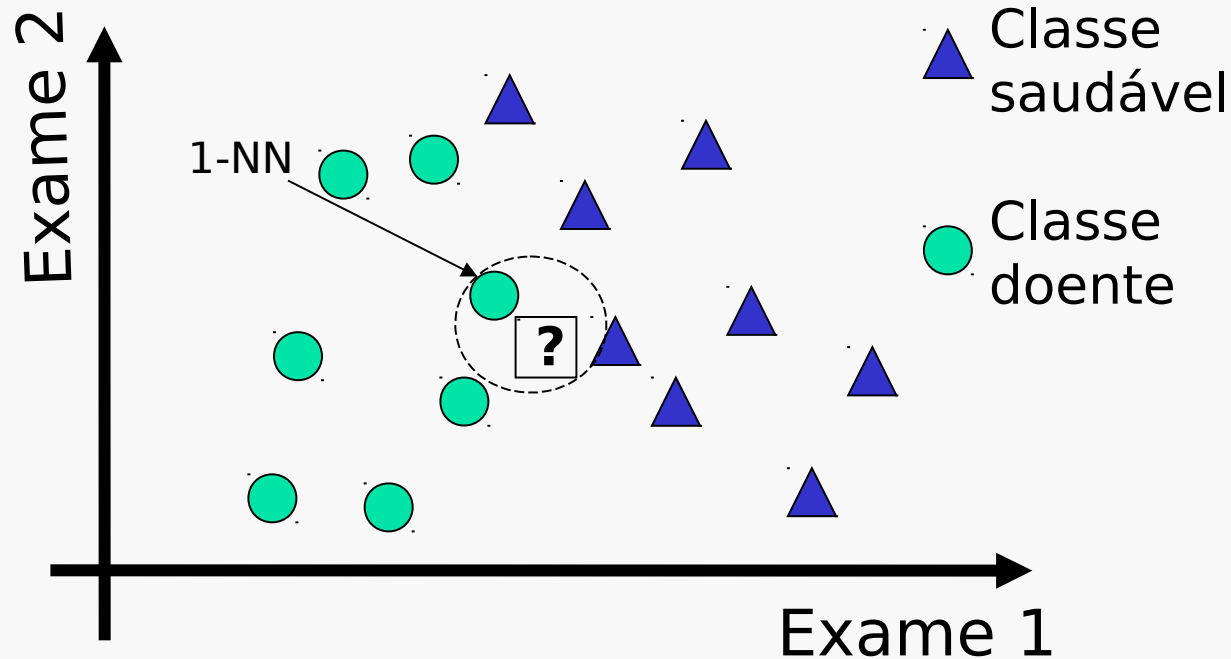
- Rotula novos objetos com base nos exemplos do conjunto de treinamento mais próximos a ele
 - É um algoritmo **preguiçoso** (*lazy*)
 - Não aprende **modelo compacto**, memoriza objetos de treinamento
 - Adia computação para a fase de classificação
 - Baseado em **informações locais**
 - Pode ser utilizado em **classificação** e **regressão**, sem necessidades de alterações significativas
 - Há variações de acordo com o **número de vizinhos** mais próximos adotado

Algoritmo 1-vizinho mais próximo

- Variação mais simples: 1-NN
 - *1-Nearest Neighbour*
 - Cada objeto representa um **ponto no espaço** de entradas
 - Definindo **métrica**, é possível **calcular distâncias**
 - Métrica mais usual: **distância euclidiana**
 - **Treinamento**: memoriza exemplos rotulados do conjunto de treinamento
 - **Classificação de novo exemplo**: classe do exemplo de treinamento mais próximo

Algoritmo 1-vizinho mais próximo

- Ex. 1-NN



Algoritmo k-vizinhos mais próximos

- Extensão imediata do 1-NN considerando mais vizinhos
 - k vizinhos mais próximos
 - k é parâmetro do algoritmo
 - Cada vizinho vota em uma classe
 - Previsões são então agregadas

Classificação

$$f^*(\mathbf{x}_t) \leftarrow \text{moda}(y_1, \dots, y_k)$$

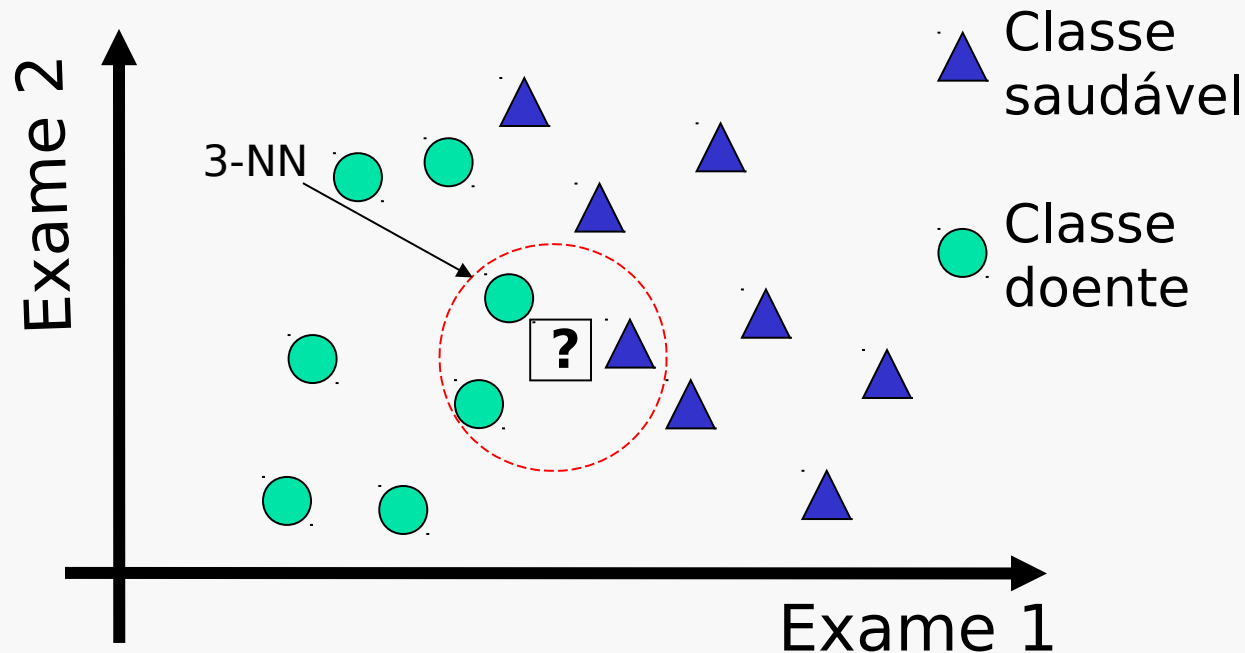
Regressão

$$f^*(\mathbf{x}_t) \leftarrow \text{média}(y_1, \dots, y_k)$$

ou $f^*(\mathbf{x}_t) \leftarrow \text{mediana}(y_1, \dots, y_k)$

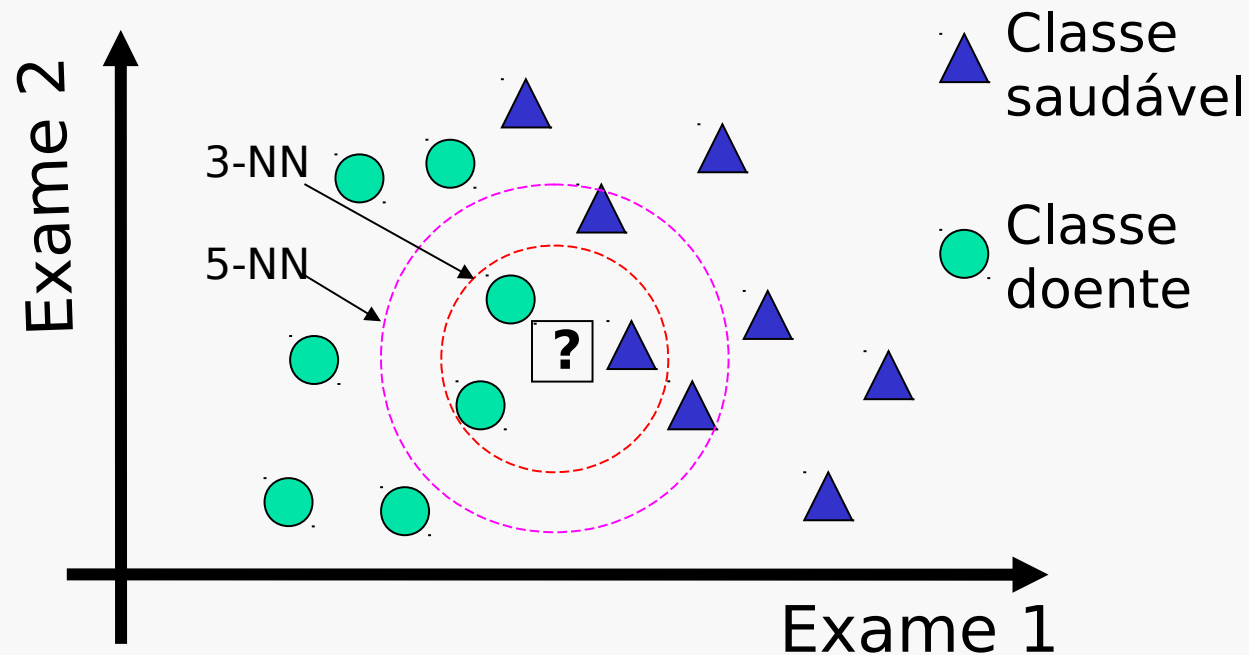
Algoritmo k-vizinhos mais próximos

■ Ex.



Algoritmo k-vizinhos mais próximos

■ Ex.



Quantos vizinhos?

- k muito grande
- k muito pequeno

Quantos vizinhos?

- k muito grande
 - Vizinhos podem ser muito diferentes
 - Predição tendenciosa para classe majoritária
 - Custo computacional mais elevado
- k muito pequeno
 - Não usar informação suficiente
 - Previsão pode ser instável

Frequentemente usa k pequeno e ímpar (3, 5, ...). Valores pares não são usuais em classificação por poderem levar a empates

k-NN: exemplo

- Supor o conjunto de dados:

sexo	cirurgia	TSH	TT4	TI	classe
f	f	5.8	156.2	f	neg
f	f	7.3	152.9	v	neg
m	v	8.8	148.4	f	hipo_prim
f	f	6.9	132.7	f	hipo_comp
m	f	5.4	150.9	f	neg

Min 5.4 132.7

Max 8.8 156.2

k-NN: exemplo

- Conjunto de dados normalizado:

sexo	cirurgia	TSH	TT4	TI	classe
f	f	0.12	1.0	f	neg
f	f	0.56	0.86	v	neg
m	v	1.00	0.67	f	hipo_prim
f	f	0.44	0.0	f	hipo_comp
m	f	0.0	0.77	f	neg

Min 0.0 0.0

Max 1.0 1.0

k-NN

- Distância heterogênea:
 - Considera atributos qualitativos e quantitativos

$$d_H(x_i, x_j) = \sqrt{\sum_{r=1}^d d'(a_r(x_i), a_r(x_j))^2}$$

- Em que:

$$d'(a, b) = \left\{ \begin{array}{l} 1, \text{ se } a \text{ e } b \text{ são qualitativos e } a \neq b \\ 0, \text{ se } a \text{ e } b \text{ são qualitativos e } a = b \\ a - b, \text{ se } a \text{ e } b \text{ são contínuos} \end{array} \right\}$$

k-NN: exemplo

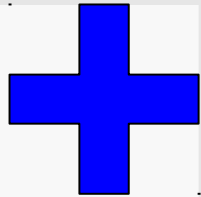
- Como classificar o exemplo com 1-NN?
 - $\mathbf{x} = \langle f \ f \ 7.0 \ 150.0 \ v \rangle$: esse paciente é o que?
 - Primeiro uso mins e maxs do treinamento para normalizar os atributos 3 e 4
 - $\mathbf{x} = \langle f \ f \ 0.47 \ 0.74 \ v \rangle$
 - Agora vejo distância em relação a cada exemplo de treino
 - $d(\mathbf{x}, \mathbf{x}_1) = 1.09$; $d(\mathbf{x}, \mathbf{x}_2) = 0.15$
 - $d(\mathbf{x}, \mathbf{x}_3) = 1.81$; $d(\mathbf{x}, \mathbf{x}_4) = 1.24$
 - $d(\mathbf{x}, \mathbf{x}_5) = 2.22$
 - Exemplo mais próximo = \mathbf{x}_2 , então classe = neg

E se o atributo da nova instância estiver for a desses limites?

k-NN: exemplo

- E se fosse 3-NN?
 - Distância em relação a cada exemplo de treino
 - $d(\mathbf{x}, \mathbf{x}_1) = 1,09$; $d(\mathbf{x}, \mathbf{x}_2) = 0,15$
 - $d(\mathbf{x}, \mathbf{x}_3) = 1,81$; $d(\mathbf{x}, \mathbf{x}_4) = 1,24$
 - $d(\mathbf{x}, \mathbf{x}_5) = 2,22$
 - Exemplos mais próximos = \mathbf{x}_2 , \mathbf{x}_1 e \mathbf{x}_4 então classe = neg (maioria)

Análise do algoritmo



- Vantagens:
 - O algoritmo de treinamento é simples
 - Armazenar os objetos
 - É aplicável mesmo em problemas complexos
 - É um algoritmo naturalmente incremental
 - Novos exemplos \Rightarrow basta armazená-los na memória

Análise do algoritmo

- Desvantagens:
 - Não obtém uma representação compacta dos dados
 - Não se tem modelo explícito a partir dos dados
 - Predição pode ser custosa
 - Requer calcular distâncias a todos os objetos de treinamento
 - É afetado pela presença de atributos redundantes e irrelevantes
 - Problemas com dimensionalidade elevada
 - Objetos ficam equidistantes

Problemas com dimensionalidade

- Considere 100 pontos distribuição uniforme:
 - Em um quadrado com lado de 1 unidade
 - Em um cubo com lado de 1 unidade
 - ...

Distância média entre dois pontos:

d	Distância média
2	0,49
3	0,65
4	0,77
5	0,88
...	
10	1,28

Distância entre dois pontos quaisquer aumenta com dimensões \Rightarrow densidade diminui e o conjunto de dados fica esparsa, rarefeito
(para 10 dimensões, a distância média é maior que o tamanho do lado do hipercubo!)

Problemas com dimensionalidade

- Considere 100 pontos distribuição uniforme:
 - Em um quadrado com lado de 1 unidade
 - Em um cubo com lado de 1 unidade
 - ...

Distância média entre dois pontos:

d	Distância média
2	0,49
3	0,65
4	0,77

Distância entre dois pontos quaisquer aumenta com dimensões \Rightarrow densidade diminui e o conjunto de dados fica esparsa, rarefeito

Uma das formas de reduzir esse impacto é selecionar subconjunto de atributos relevantes: área de pesquisa

Seleção de Atributos

cia
do

Referências

- Slides de Ana Carolina Lorena
- Capítulo 1 do livro Inteligência Artificial: uma abordagem de Aprendizado de Máquina, 2011
- Alguns slides foram baseados em apresentações de:
 - Prof Dr André C. P. L. F. Carvalho
 - Prof Ricardo Campello
 - Profa Solange O. Rezende
 - Prof Dr Marcilio C. P. Souto
 - Prof Dr Carlos Soares