

Aprendizado de Máquina

Aula 3: Tarefas de aprendizado (parte 1)

André C. P. L. F de Carvalho
ICMC/USP

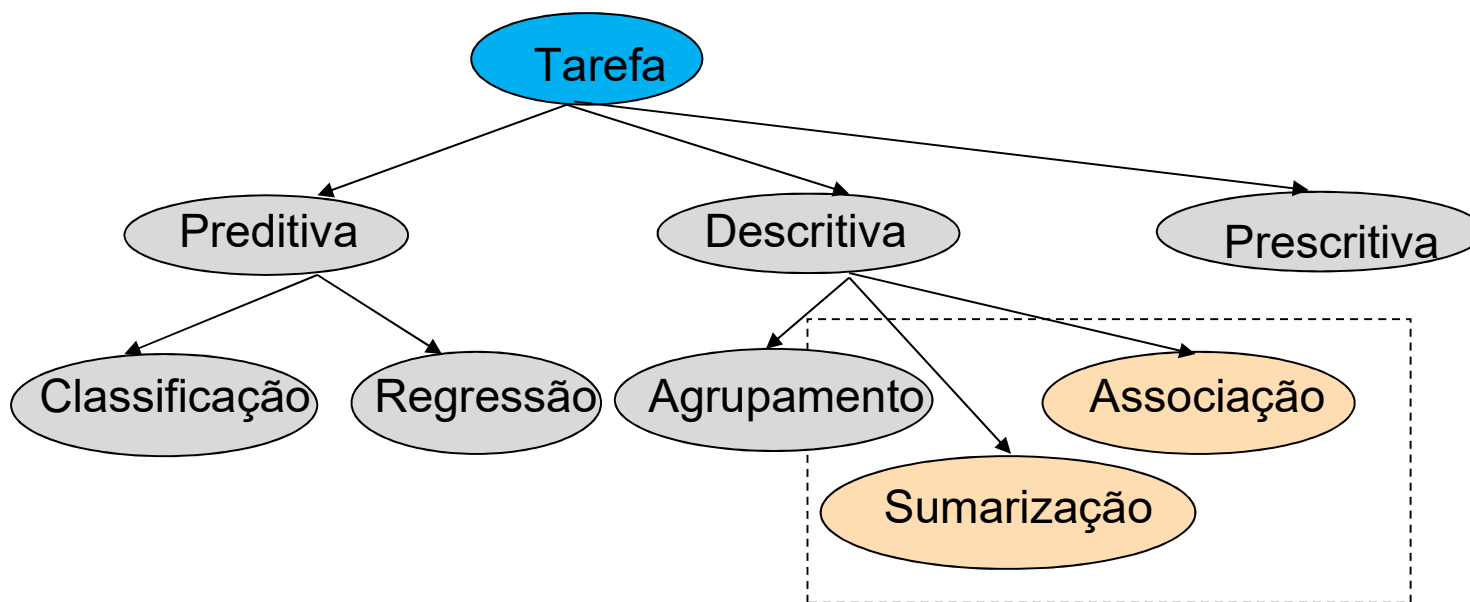
andre@icmc.usp.br



Tópicos a serem cobertos

- Principais tarefas de aprendizado
- Tarefas preditivas
- Tarefas descritivas
- Tarefas prescritivas
- Exemplos

Tarefas de aprendizado



Tarefas preditivas

- Tarefas que precisam de um modelo capaz de prever o rótulo (atributo alvo) de seus exemplos
 - A partir do valor de cada um dos atributo preditivos do exemplo
 - Modelo preditivo
- Para a induzir o modelo pode ser usado um algoritmo de aprendizado de máquina
 - Algoritmo ensina (treina) o modelo a desempenhar bem sua tarefa por meio de um processo de aprendizado
 - Para isso, usa um conjunto de dados de treinamento
 - Um outro conjunto de dados, conjunto de dados de teste, avalia o quão bem o modelo aprendeu a realizar a tarefa

Tarefa de regressão

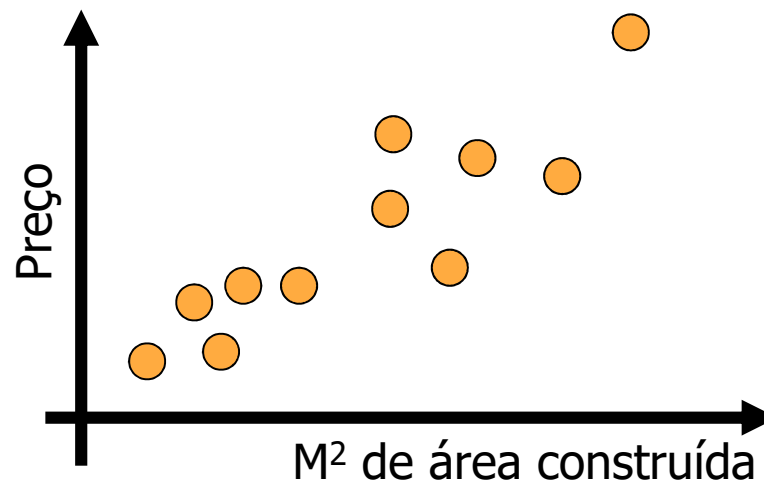
- Objetivo: aprender uma função capaz de associar a descrição de um exemplo a um valor real
 - Aproximação de função
- Exemplos:
 - Prever valor de mercado de um imóvel
 - Prever o lucro de um empréstimo bancário
 - Prever tempo de internação de um paciente
 - Prever que nota alguém vai tirar em uma prova

Tarefa de regressão

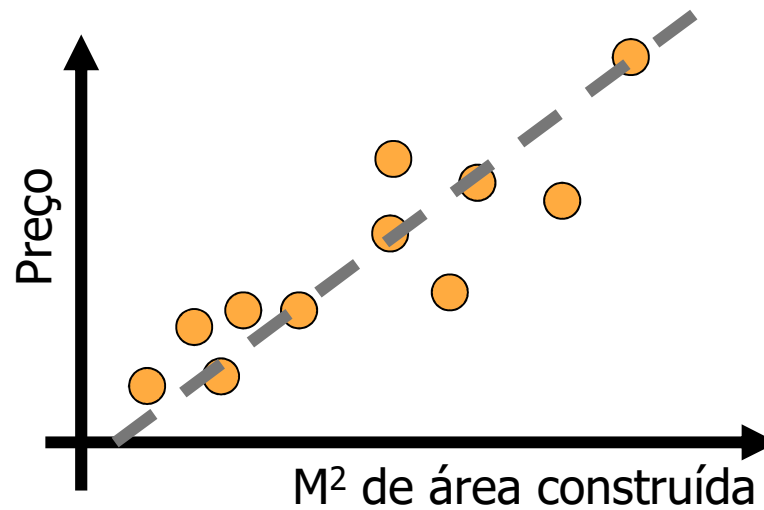
- Imobiliária vendenada
 - Um corretor é o especialista em dar preços
 - Já vendeu várias casas
 - Usa uma ideia simples para estimar valor de uma casa:
 - Preço é igual a 10.000 vezes o número de minutos que demora para percorrer toda a casa
 - Outro corretor acha que pode estimar parecidos (ou melhor) o valor usando área construída

Tarefa de regressão

Aproximação de função



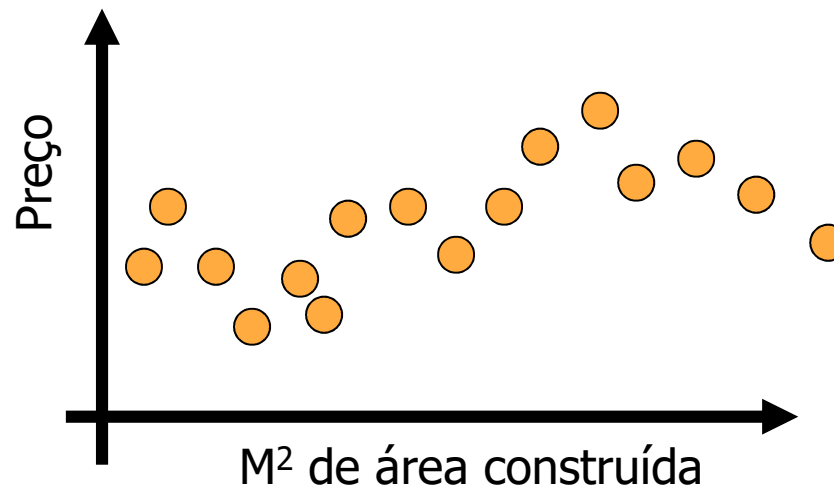
Tarefa de regressão



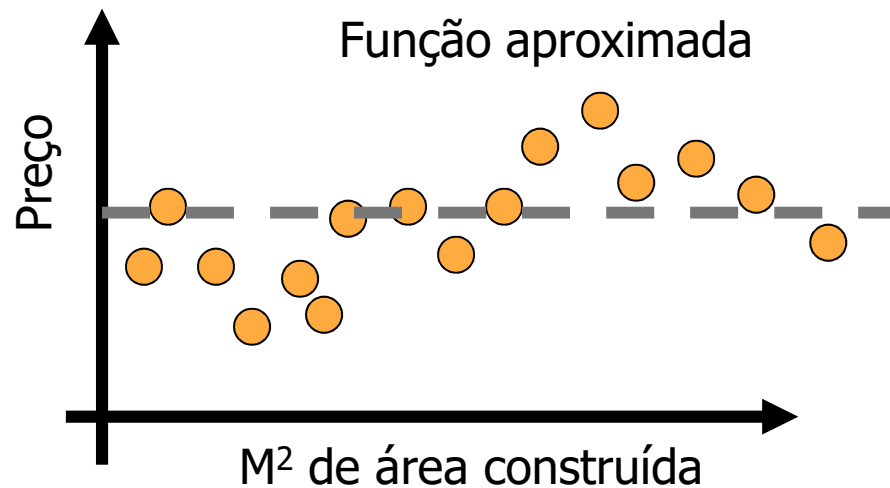
Tarefa de regressão

- O vendedor se saiu tão melhor na estimativa, que uma imobiliária maior ofereceu um emprego para ele
 - Ganhando muito mais, é claro
 - Como o uso de uma função linear gerou boas previsões, ele nem se preocupou em procurar outra função

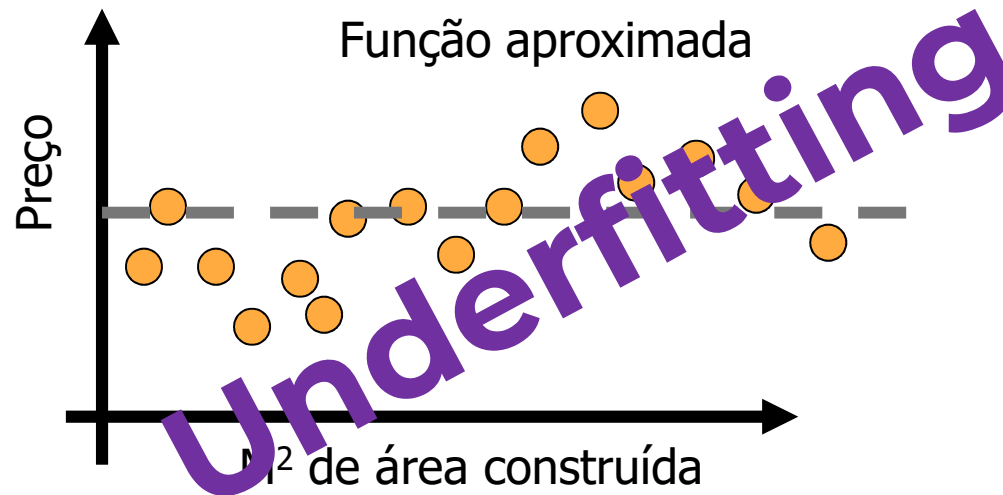
Tarefa de regressão



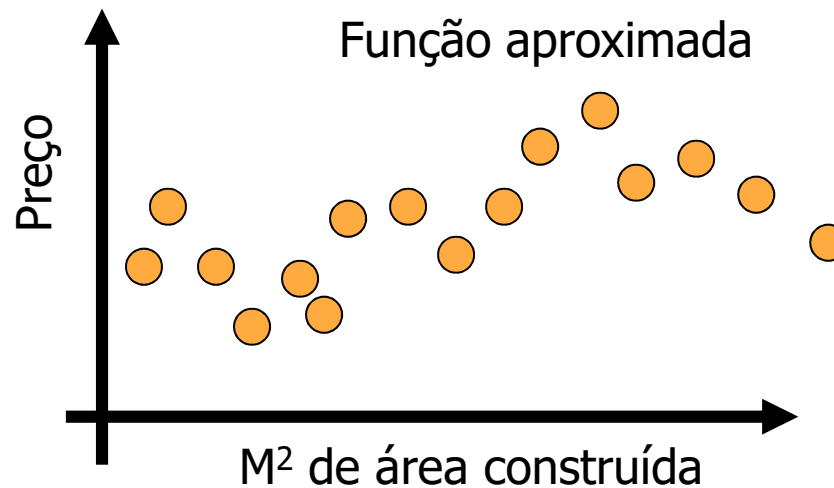
Na vida real, nada é tão simples...



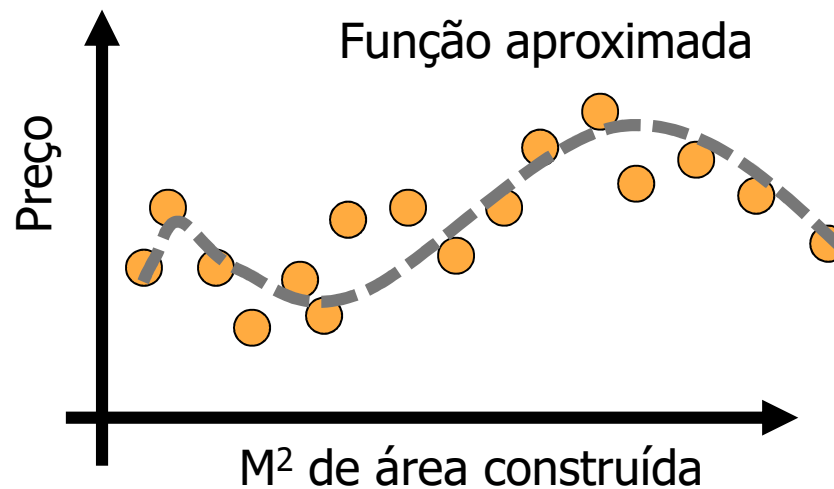
Na vida real, nada é tão simples...



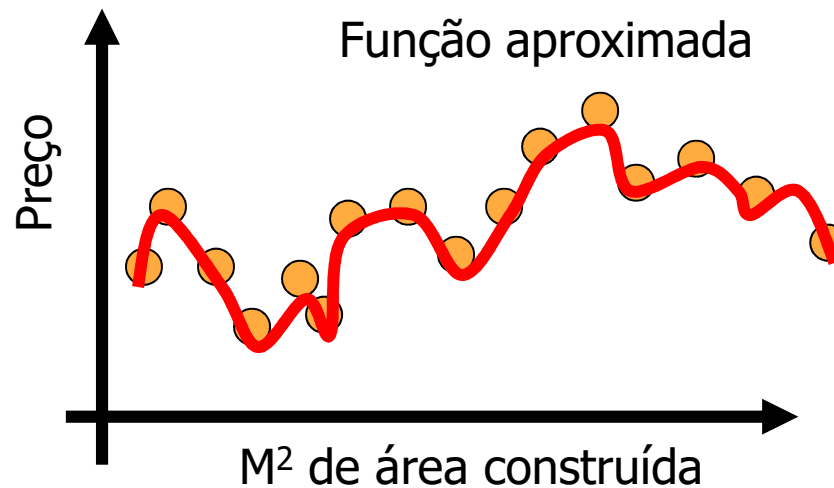
Na vida real, nada é tão simples...



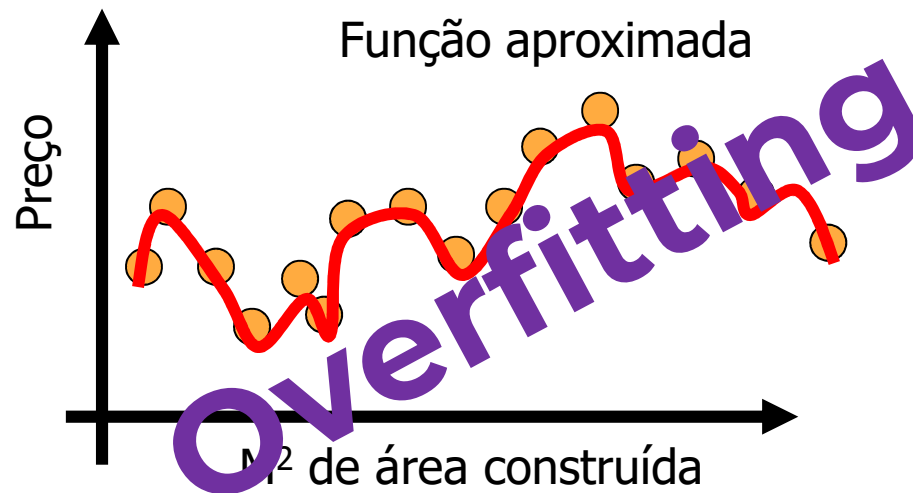
Na vida real, nada é tão simples...



Na vida real, nada é tão simples...



Na vida real, nada é tão simples...



Tarefa de classificação

- Objetivo: aprender função que associa descrição de um objeto a sua classe
 - Fronteira de decisão
- Exemplos:
 - Definir a função de uma proteína
 - Diagnosticar um paciente como tendo ou não uma determinada doença
 - Decidir se um sinistro foi fraudulento

Tarefa de classificação

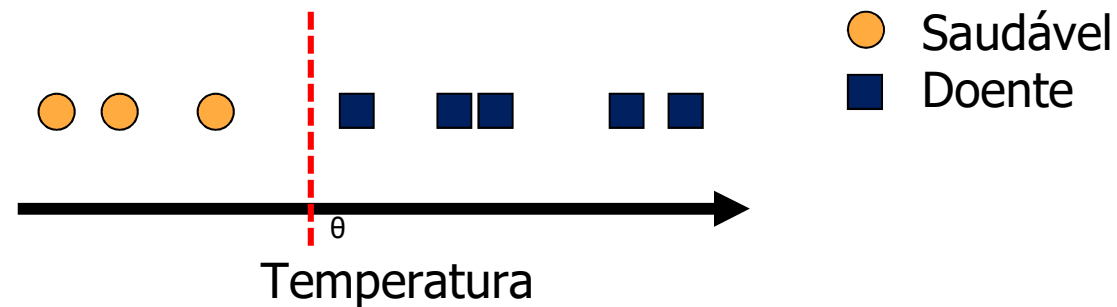
- Posto médico Daquiproceu
 - Tem um arquivo com o histórico de vários atendimentos e diagnósticos
 - Um paciente, ao sentir alguns sintomas, vai ao posto para uma consulta médica
 - O único médico, faltou
 - Mas um aluno de medicina, estagiário, pode anotar os sintomas
 - Será que o estagiário fazer um bom pré-diagnóstico?

Tarefa de classificação

- Sintomas coletados pelo estagiário:
 - Temperatura

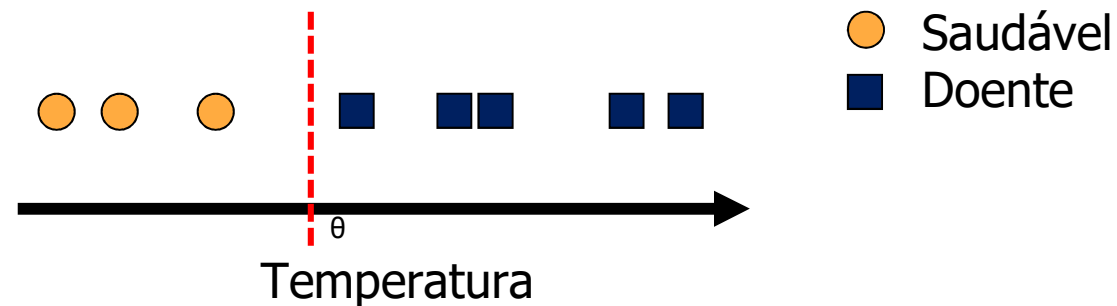
Tarefa de classificação

- Forma mais simples de resolver



Tarefa de classificação

- Forma mais simples de resolver



Função estimada: diagnóstico = $f(\text{temperatura})$

Se temperatura $> \theta$

Então doente

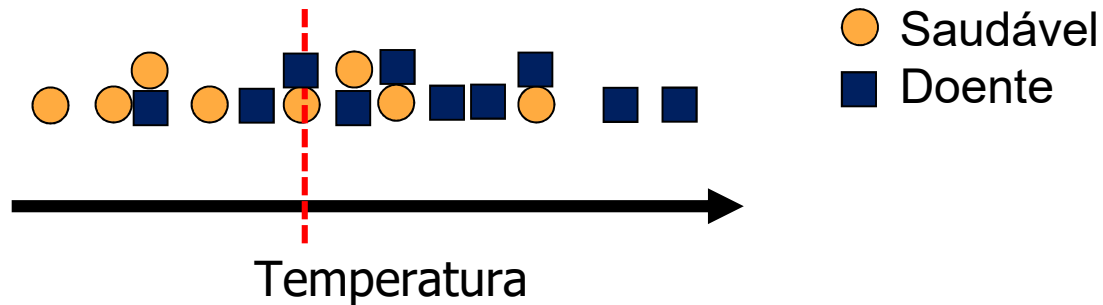
Senão saudável

Tarefa de classificação

- Basta encontrar um valor (limiar) de temperatura que separa
 - Doentes
 - Saudáveis
- Mas todo problema de classificação é tão simples assim?
 - Uso apenas da temperatura gera um bom modelo preditivo?

Tarefa não é tão simples

- Supor a inclusão de outros pacientes



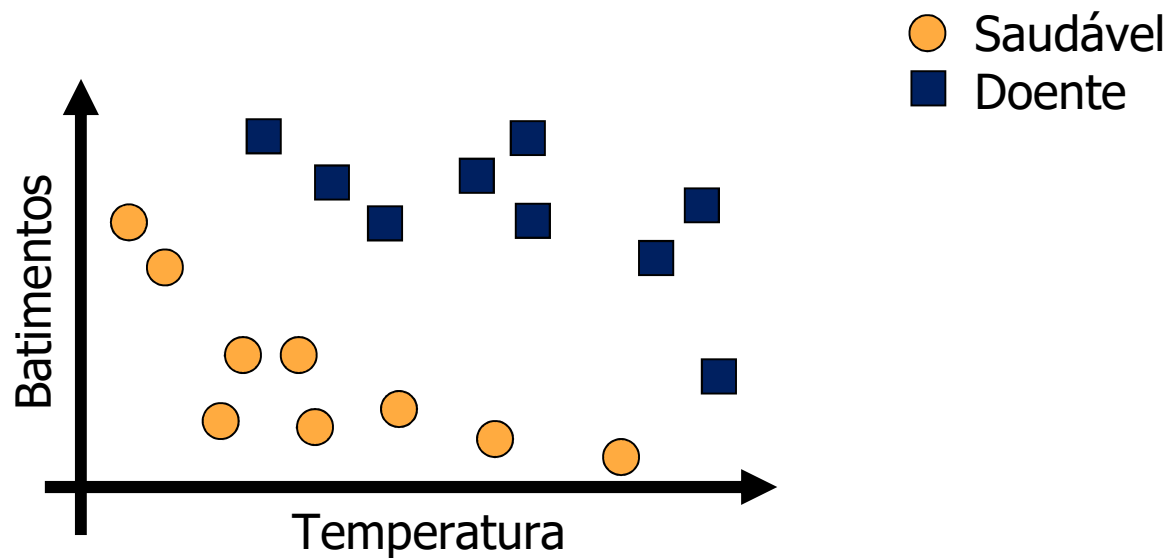
- Alternativa: considerar outros sintomas para o diagnóstico

Tarefa não é tão simples

- Enfermeira decide coletar outro sintoma:
 - Taxa de batimentos cardíacos
- Agora, são coletados do paciente:
 - Taxa de batimentos cardíacos
 - Temperatura

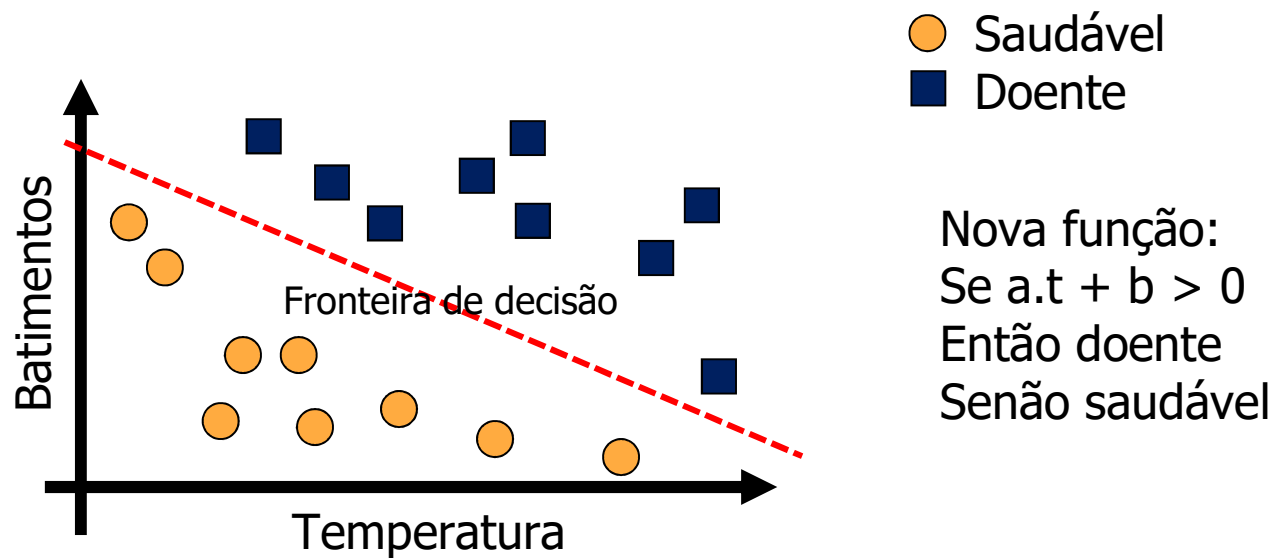
Tarefa não é tão simples

- Incluindo taxa de batimentos cardíacos



Tarefa de classificação

- Função linear permite um bom diagnóstico

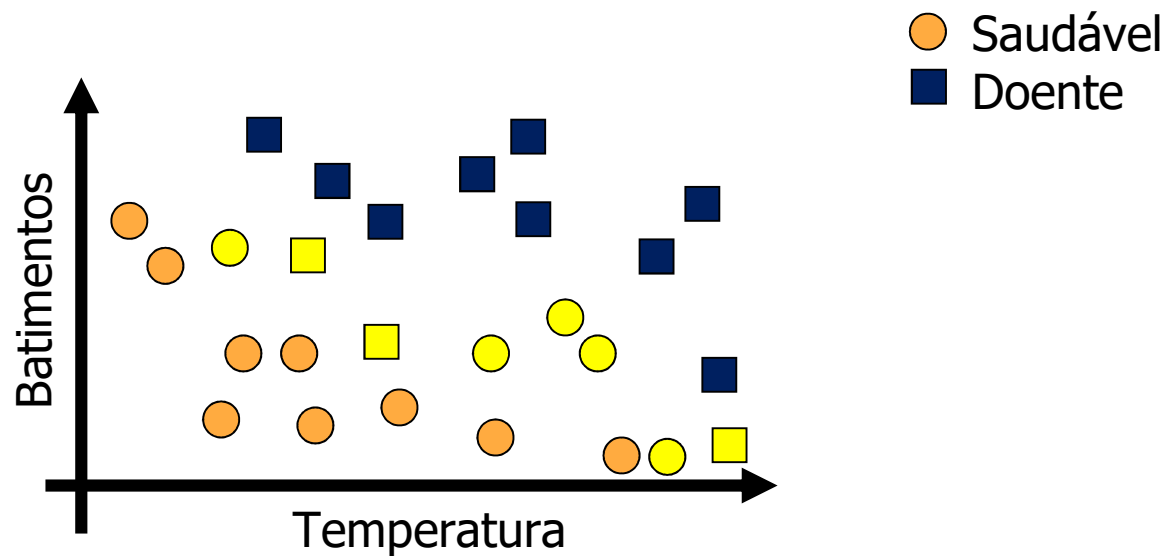


Tarefa de classificação

- Basta encontrar uma função linear que separa pacientes doentes de saudáveis
 - Inclinação da reta e ponto onde cruza o eixo da ordenada
- Espaço de pacientes
 - Ordenada: taxa de batimentos cardíacos
 - Abscissa: temperatura
- Mas toda tarefa de classificação é simples assim?

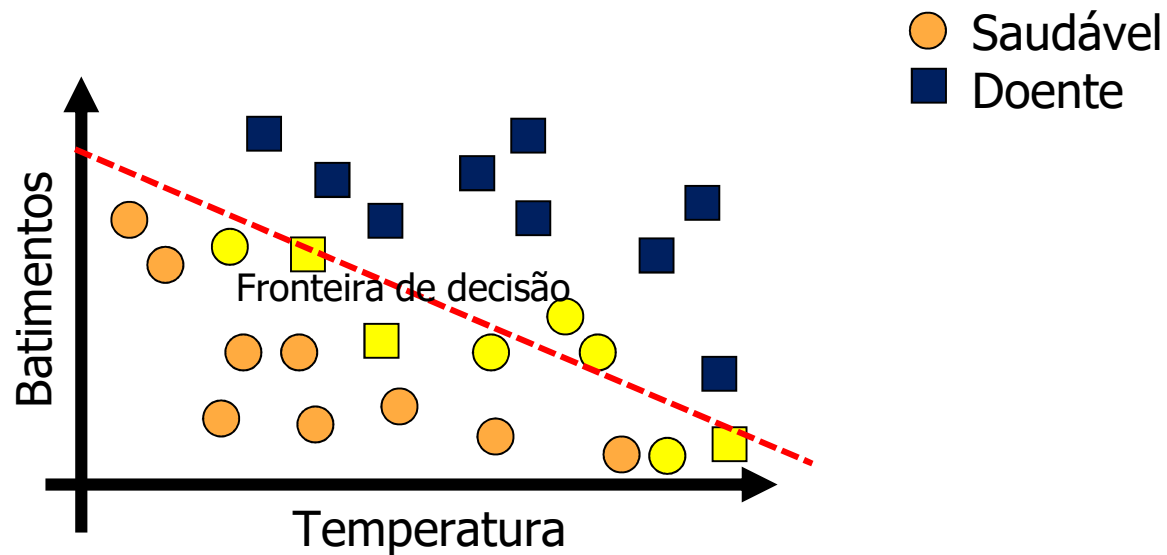
Tarefa não é tão simples

- Supor que precisava incluir dados de outros pacientes



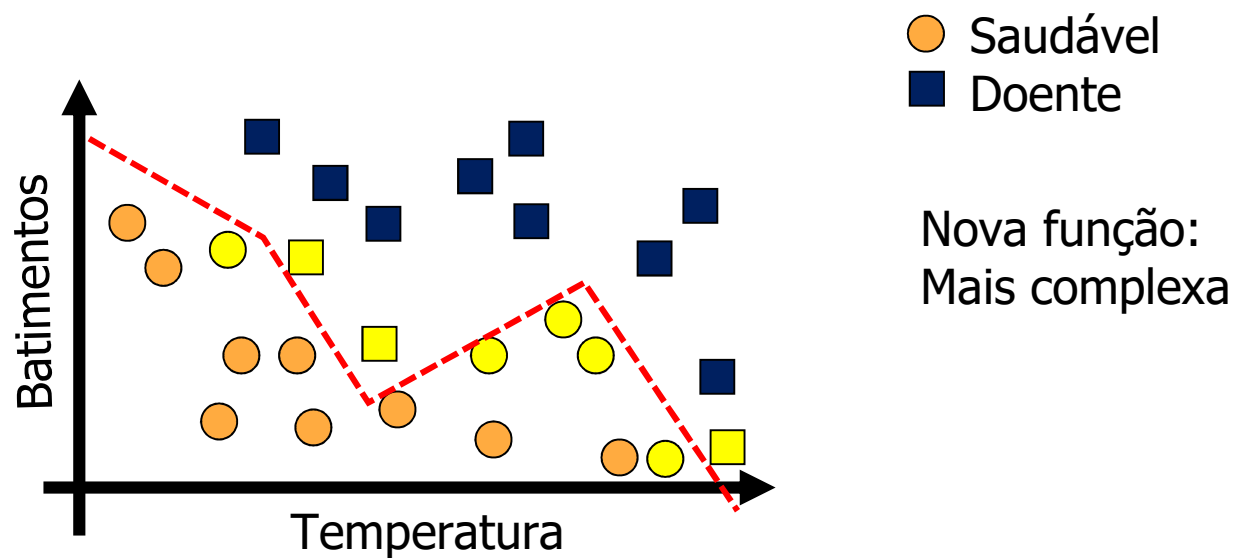
Tarefa não é tão simples

- Função linear agora não permite um bom diagnóstico



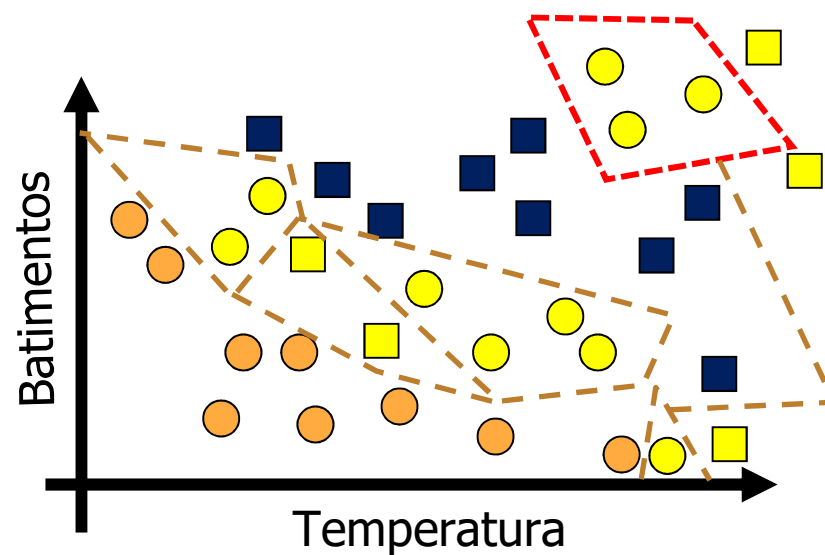
Tarefa não é tão simples

- Função não linear permite um melhor diagnóstico



Tarefa não é tão simples

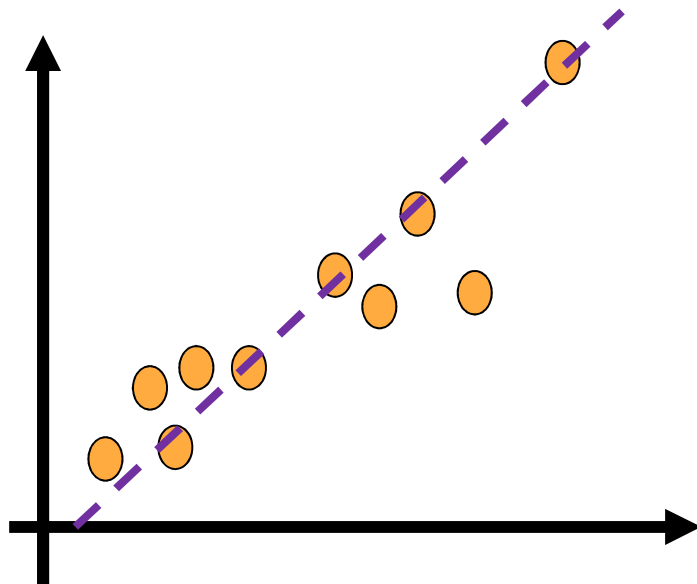
- Supor a inclusão de mais pacientes



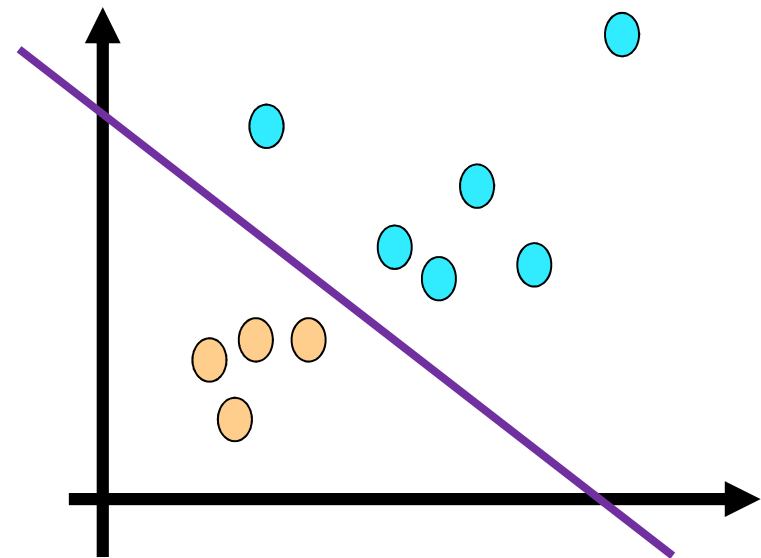
● Saudável
■ Doente

Nova função:
Muito mais complexa
Pode apresentar overfitting

Classificação vs regressão



Regressão



Classificação

Continua no próximo vídeo e
conjunto de slides

Aprendizado de Máquina

Aula 3: Tarefas de aprendizado (parte 2)

André C. P. L. F de Carvalho
ICMC/USP
andre@icmc.usp.br



Tarefas descritivas

- Também buscam por modelos em um processo de treinamento
 - Descrevem ou sumarizam dados de uma tarefa
 - Treinamento utiliza todo o conjunto de dados
 - Ex.: Agrupamento de dados
- Algumas tarefas descritivas não possuem uma fase de treinamento
 - Ex.: Tarefas de sumarização e de associação de itens frequentes

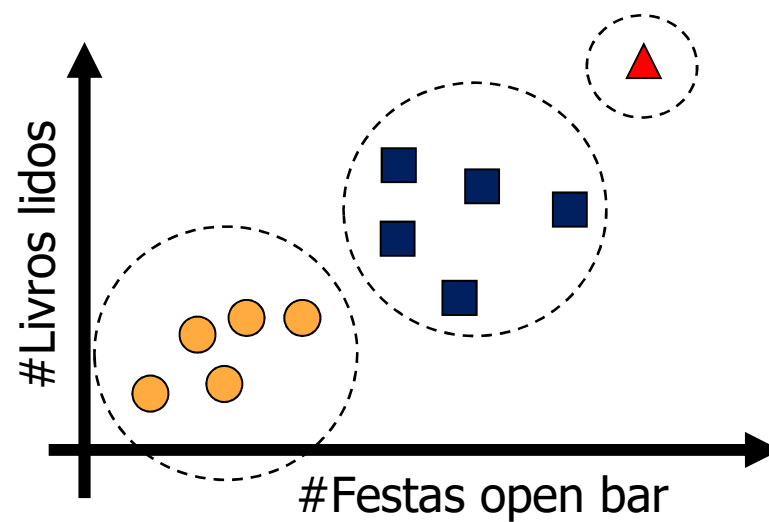
Agrupamento (Clustering)

- Objetivo: organizar objetos não rotulados em grupos (clusters)
 - De acordo com uma medida de proximidade entre objetos
- Não existe conhecimento anterior sobre:
 - Número de grupos (maioria das vezes)
 - Significado dos grupos (pode ter vários)

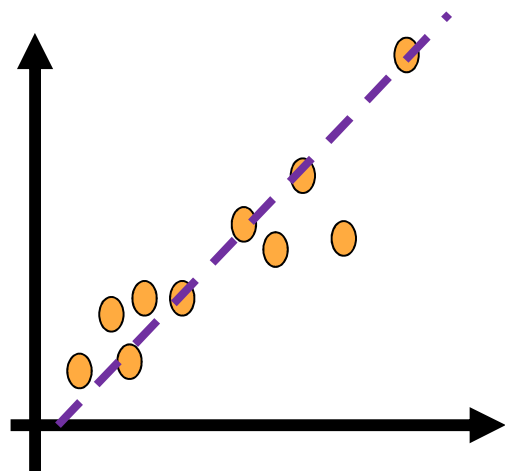
Agrupamento

- Colégio Nãovaidar
 - Tem um grande número de alunos
 - Gostaria que eles formassem grupos que possuem interesses parecidos
 - Só sabe duas coisas de cada aluno
 - Quantos livros leu no ano passado
 - Quantas vezes foi para uma festa open bar no ano passado
 - É possível sugerir bons grupos?

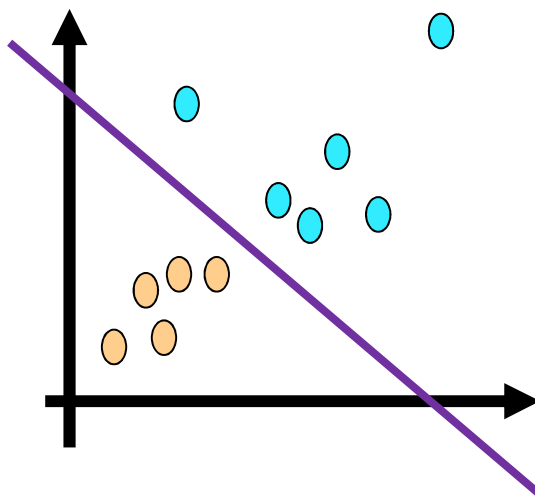
Agrupamento



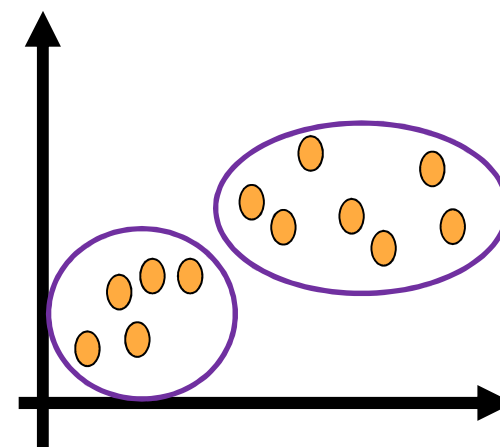
Comparação de tarefas



Regressão



Classificação



Agrupamento

Sumarização

- Objetivo: encontrar descrição simples e resumida para um conjunto de dados
- Frequentemente utilizada para:
 - Exploração interativa de dados
 - Geração automática de relatórios
 - Exemplo:
 - Definir perfis de pacientes com comorbidade

Sumarização

Escolaridade	Pressão alta	Sexo	Idade	Comorbidade
Médio	Sim	M	34	Não
Superior	Não	F	40	Não
Superior	Não	F	31	Não
Fundamental	Sim	F	18	Não
Médio	Não	M	76	Sim
Superior	Não	F	35	Não
Fundamental	Sim	M	20	Não
Superior	Não	M	76	Sim
Fundamental	Não	M	43	Não
Médio	Não	F	27	Sim

Sumarização

Escolaridade	Pressão alta	Sexo	Idade	Comorbidade
Médio	Sim	M	34	Não
Superior	Não	F	40	Não
Superior	Não	F	31	Não
Fundamental	Sim	F	18	Não
Médio	Não	M	76	Sim
Superior	Não	F	35	Não
Fundamental	Sim	M	20	Não
Superior	Não	M	76	Sim
Fundamental	Não	M	43	Não
Médio	Não	F	27	Sim

Escolaridade mais comum: Superior
 Frequência de comorbidade: 30%
 Idade média: 40
 Igualdade de sexo: S
 Menor idade: 18

Associação de itens frequentes

- Frequent itemsets
- Objetivo: dado um conjunto de itens e uma base de dados de transações
 - Encontrar conjunto de regras que, nas várias transações realizadas, associem a presença de um item à presença de outros itens
 - Conjunto de regras de associação
- Exemplo:
 - Procurar por itens que são frequentemente comprados juntos em um supermercado
 - Problema das cestas de compras

Associação de itens frequentes

- Problema das cestas de compras
 - Conjunto de transações, em que cada transação é uma compra feita em um dado supermercado

Transação	Itens comprados
1	pão, queijo, manteiga, massa
2	pão, geleia, suco
3	queijo, arroz, massa
4	queijo, vinho
5	massa, queijo, pão

Associação de itens frequentes

- Itemset: uma coleção de um ou mais itens
 - Quando possui k itens, é chamado de k-itemset
 - Ex.: {pão, geleia, suco} é um 3-itemset
- Cobertura: a fração das transações em que um itemset aparece
 - Ex.: cobertura {pão, queijo} = $2/5$ (aparece em 2 das 5 transações) = 40%
- Itemset frequente: um itemset cuja cobertura é maior ou igual a um dado threshold (limiar)
- Regra de associação: regra em que o antecedente e o consequente são itemsets
 - Ex.: se compra pão, então compra queijo

Transação	Itens comprados
1	pão, queijo, manteiga, massa
2	pão, geleia, suco
3	queijo, arroz, massa
4	queijo, vinho
5	massa, queijo, pão

Associação de itens frequentes

- Problema das cestas de compras
 - Conjunto de transações, em que cada transação é uma compra feita em um dado supermercado

Transação	Itens comprados
1	pão, queijo, manteiga, massa
2	pão, geleia, suco
3	queijo, arroz, massa
4	queijo, vinho
5	massa, queijo, pão

66% dos clientes que compraram pão também compraram queijo

75% dos clientes que compraram queijo também compraram massa

Tarefa prescritiva

- Prescreve que entrada é necessária para gerar uma dada saída
 - Contrário do aprendizado preditivo
 - Ao invés de prever o que vai acontecer, sugerir o que fazer para que algo aconteça
 - Exemplo
 - Controle de robôs
 - Gera entrada de controle para que um sistema siga uma trajetória especificada por um modelo de referência
 - Controle de processos químicos

Alchemist

- Desenvolvido em parceria com o departamento de Engenharia de Materiais da UFSCar
 - Para criar vidros novos, com uma ou mais propriedades
 - Pode prever que combinação de elementos químicos (que átomos e em que quantidade) pode gerar um vidro com uma dada propriedade
 - Experimentos iniciais: criar vidros com um dado valor de T_g
 - Relacionado a temperatura em que um composto químico se torna um vidro

Alchemist

Alchemist
Glass G. Visualization Options Help Exit

Run

Search space limitation:

	min	max
Ag20	0.0	0.0
Al203	0.0	1.0
As203	0.0	1.0
As205	0.0	1.0
B203	0.0	1.0
BaO	0.0	0.0

Methods:
Method 1

TG:
1000

Amount:
10

Results:

	Ag20	Al203	As203	As205	B203	BaO
1	0.0	0.04628...	0.15484...	0.0	0.01469...	0.0
2	0.0	0.50993...	0.00843...	0.06804...	0.0	0.0
3	0.0	0.10009...	0.14755...	0.12093...	0.01370...	0.0
4	0.0	0.07267...	0.11643...	0.22944...	0.0	0.0
5	0.0	0.62906...	0.0	0.03379...	0.11423...	0.0
6	0.0	0.79215...	0.0	0.0	0.27780...	0.0
7	0.0	0.73452...	0.00769...	0.0	0.25039...	0.0
8	0.0	0.15635...	0.38182...	0.0	0.01172...	0.0
9	0.0	0.94743...	0.04295...	0.0	0.13999...	0.0
10	0.0	0.21767...	0.04479...	0.36112...	0.0	0.0

Clan All Save Discard

Alchemist
Glass G. Visualization Options Help Exit

Run

Search space limitation:

	min	max
Ag20	0.0	0.0
Al203	0.0	1.0
As203	0.0	1.0
As205	0.0	1.0
B203	0.0	1.0
BaO	0.0	0.0

Methods:
Method 1

TG:
1000

Amount:
10

Results:

	W03	Y203	Yb203	ZnO	ZrO2	TG
1	0.0	0.0	0.0	0.0	0.0	992.771...
2	0.0	0.0	0.0	0.0	0.0	992.999...
3	0.0	0.0	0.0	0.0	0.0	990.845...
4	0.0	0.0	0.0	0.0	0.0	993.497...
5	0.0	0.0	0.0	0.0	0.0	995.338...
6	0.0	0.0	0.0	0.0	0.0	995.340...
7	0.0	0.0	0.0	0.0	0.0	990.269...
8	0.0	0.0	0.0	0.0	0.0	1000.29...
9	0.0	0.0	0.0	0.0	0.0	1009.44...
10	0.0	0.0	0.0	0.0	0.0	990.823...

Clan All Save Discard

Fim da
apresentação