



CLASSES DE PROBLEMAS

André Gustavo Adami
Daniel Luis Notari

INTRODUÇÃO

A classe do problema é baseada na forma de como a aprendizagem é realizada (supervisionada, não-supervisionada ou por reforço)

A classe do problema define quais algoritmos são aplicáveis

Trataremos somente das classes de problemas com **aprendizagem supervisionada e não supervisionada** por demandarem dados para a aprendizagem

CLASSES DE PROBLEMAS

- **Aprendizagem Supervisionada**

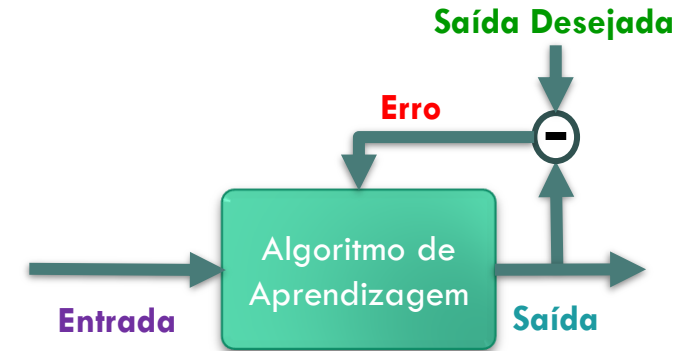
Problemas que demandam realizar previsões com base em um conjunto de exemplos (experiências)

O conjunto de exemplos é formado por dados do domínio do problema e a saída desejada e o resultado da aprendizagem é mensurável

- **Aprendizagem Não-Supervisionada**

Problemas que não possuem um conjunto definido de categorias ou classes, mas deseja-se organizar ou descobrir padrões a partir dos dados

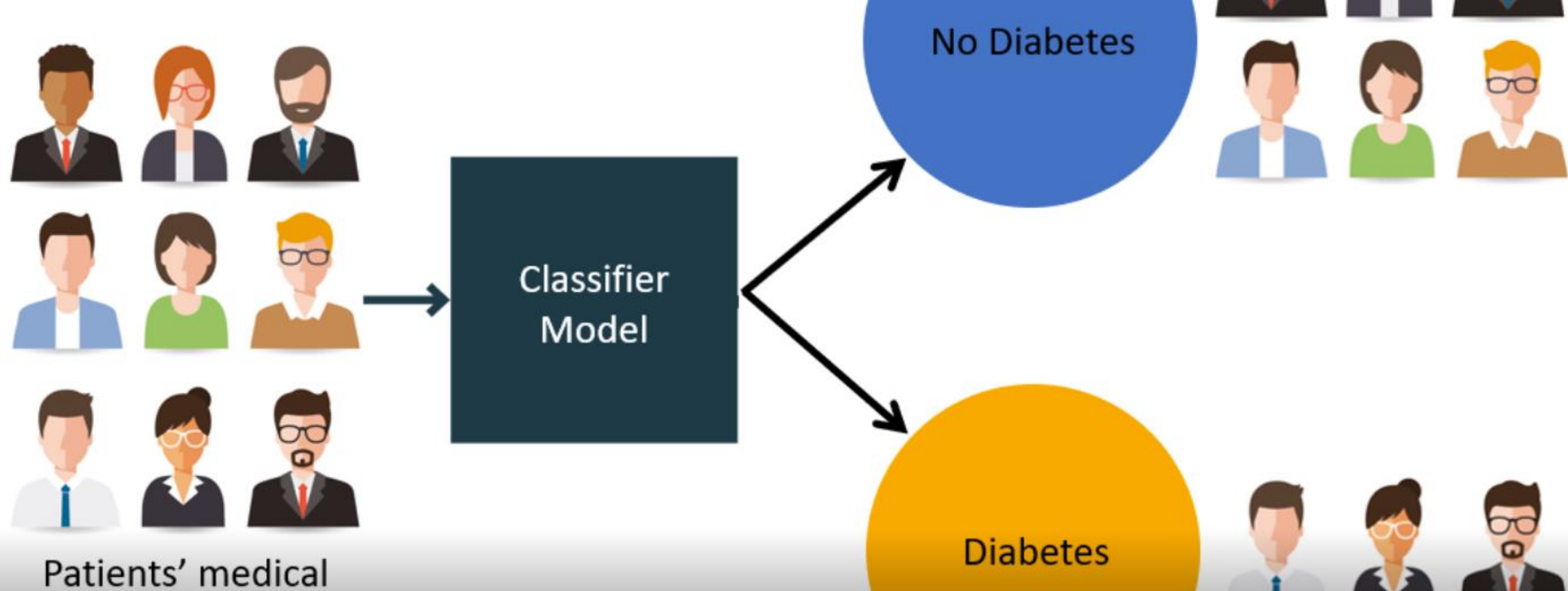
Como não possui a saída desejada, o resultado da aprendizagem pode ser qualitativo ou demanda especialista do domínio



CLASSES DE PROBLEMAS

As classes de problemas podem ser refinadas em cada tipo de aprendizagem



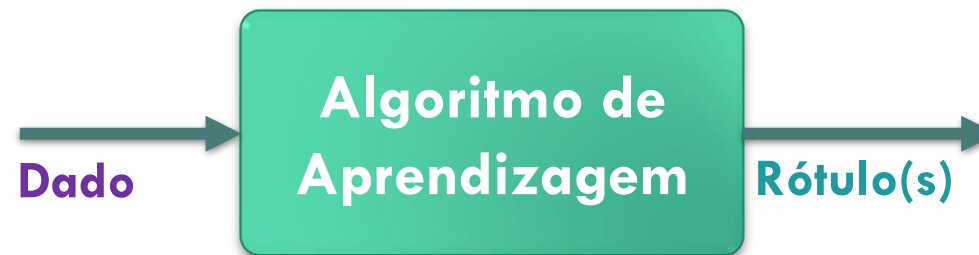


CLASSIFICAÇÃO

Classes de Problemas
Supervisionado

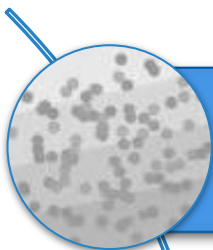
CLASSIFICAÇÃO

É o processo de prever/atribuir uma categoria/classe (saída é uma variável discreta/categórica, também conhecido por **rótulo**) de um determinado domínio a uma dada amostra



Exemplos de classificação incluem determinar se uma imagem é de um gato, cachorro ou cavalo, se uma pessoa está em um grupo de alto risco para uma determinada doença, se uma música é rock, clássica ou samba (ou uma mistura...)

CLASSIFICAÇÃO: TIPOS



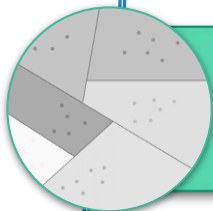
Unária

Identificar um única classe



Binária

Escolher 1 classe de um conjunto de 2 classes



Multi-classe

Escolher 1 classe de um conjunto de N classes



Multi-rótulo

Escolher M classes de um conjunto de N classes

CLASSIFICAÇÃO UNÁRIA

Refere-se aos problemas que possuem somente dados da classe de interesse (ou alvo)

- Conhecida também como **one-class classification**

A classe negativa (o complemento do espaço de classes) pode não ter dados disponíveis, dados amostrados de forma insuficiente, ou não ser bem definida



Normal Chair Objects ●



Abnormal Chair Objects ◆

CLASSIFICAÇÃO UNÁRIA: APLICAÇÕES

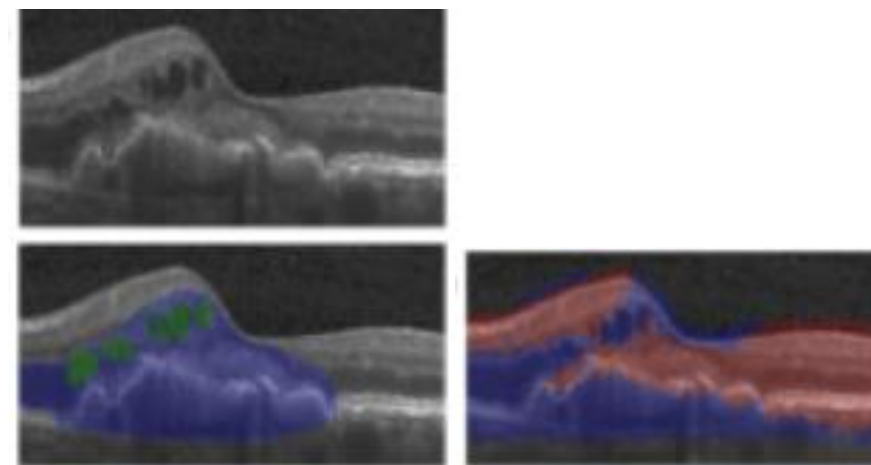
Detecção de Anomalias

- Em problemas de monitoramento de falha de máquinas, é muito fácil obter medidas das máquinas em operação normal
- Isso não é verdade para as falhas (mesmo que obter medidas de tais situações envolve aguardar muito tempo, custos financeiros altos ou até um risco para as pessoas)



Diagnóstico de Doenças

- É fácil obter dados de pacientes que tiveram algum tipo de doença comum
- Entretanto, é muito difícil ter dados que não tiveram tal doença porque nunca foram testados (custo do teste é proibitivo)



CLASSIFICAÇÃO BINÁRIA

Refere-se aos problemas que possuem somente 2 classes a serem discriminadas

- Conhecida também como problema de **detecção** (originário do problema de radares/sonares)

Existem diversos algoritmos de aprendizado de máquina desenvolvidos para tal classe

Vasta disponibilidade de material para avaliação de desempenho (acurácia, precisão, taxa de erro, falso alarme, entre outros)



CLASSIFICAÇÃO BINÁRIA



Uma das classes de problemas mais pesquisadas devido a sua grande aplicabilidade ao mundo real conjugado com a disponibilidade de algoritmos e metodologias

CLASSIFICAÇÃO MULTICLASSE

Refere-se aos problemas que possuem mais de 2 classes (mutuamente exclusivas)

- Também conhecida por **reconhecimento** (por exemplo, reconhecimento de imagens) ou **identificação** (por exemplo, identificação de pessoas utilizado atividades físicas)

A saída do sistema continua ainda sendo uma única classe do conjunto

Nem todos os algoritmos de aprendizagem e medidas de desempenho são inerentemente multiclasse





“

...there are **known knowns**;
there are things we know we know.
We also know there are **known
unknowns**; that is to say, we know
there are some things we do not
know. But there are also **unknown
unknowns** – the ones we don't
know we don't know.

”

Donald Rumsfeld, 2002

Secretário de Defesa Americano (2001-2006)

Explicação das limitações dos relatórios da inteligência

<https://www.youtube.com/watch?v=REWeBzGuzCc>

CLASSIFICAÇÃO MULTICLASSE: + DEFINIÇÕES

CONJUNTO FECHADO



Os conjuntos de dados utilizados para o aprendizado e a avaliação deste aprendizado vêm das mesmas classes (classes do domínio do problema são conhecidas - **known knowns & known unknowns**)

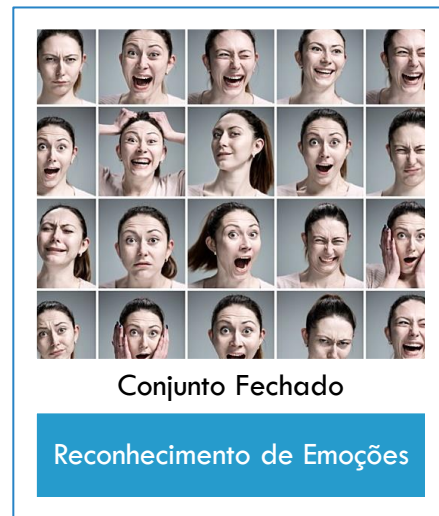
Maioria dos modelos de aprendizado de máquina atuais

CONJUNTO ABERTO



Os dados utilizados para o aprendizado vêm de um subconjunto de classes do problema (i.e., existem classes do domínio do problema que não foram aprendidas – **unknown unknowns**)

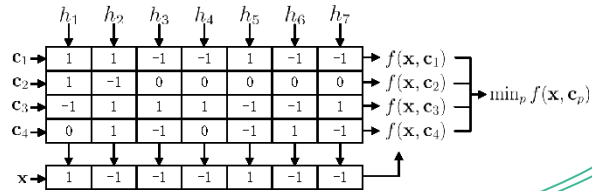
CLASSIFICAÇÃO MULTICLASSE



CLASSIFICAÇÃO MULTICLASSE: + ABORDAGENS

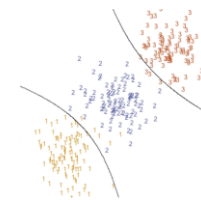
Error Correcting Output Coding

O problema é convertido em um grande número de classificadores binários. A combinação dos classificadores cria um código (code word) binário (2 classes) que é comparado ao que é produzido pelos classificadores. A classe é definida como a mais semelhante da saída esperada para a classe.



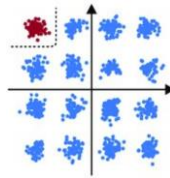
Extensões dos Classificadores Binários

Alguns classificadores binários são estendidos para lidar com problemas de classificação multi-classe



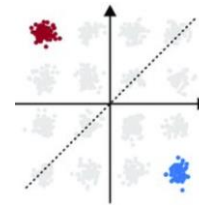
One vs Rest

Um problema de k classes é dividido em k problemas binários (conhecidos como One vs All)



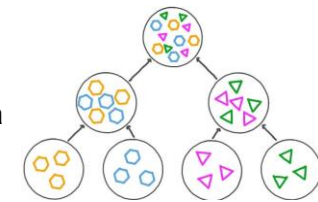
One vs One

As k classes são combinadas duas a duas produzindo $k(k-1)/2$ classificadores binários (conhecidos como All pairs ou All vs All)



Classificador Hierárquico

As classes são divididas em uma árvore de tal maneira que cada nível realiza um refinamento da classificação

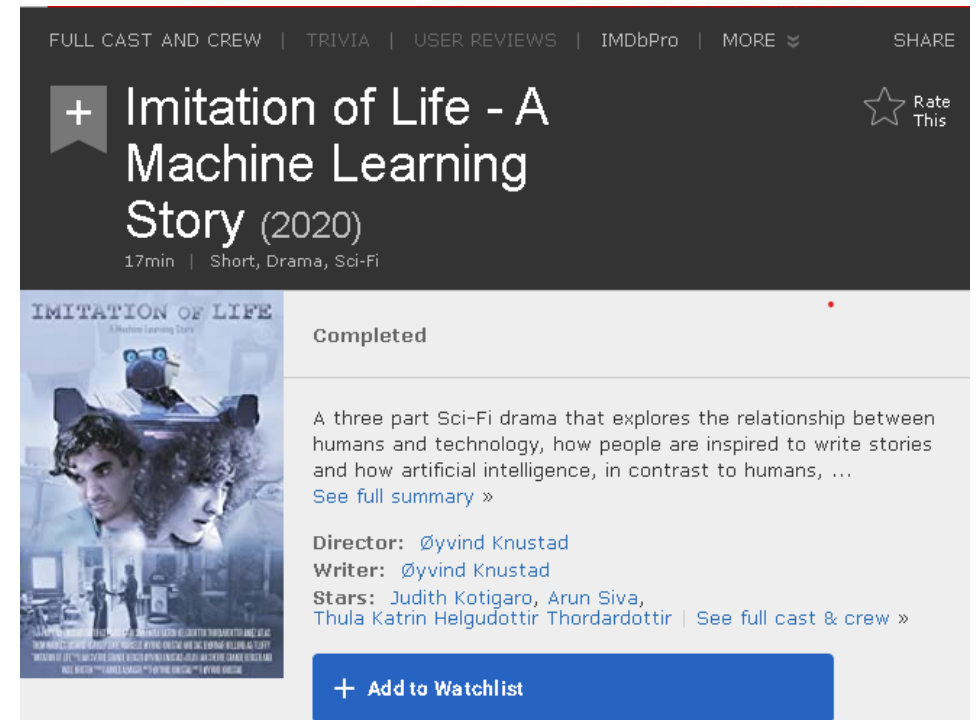


CLASSIFICAÇÃO MULTI-RÓTULO

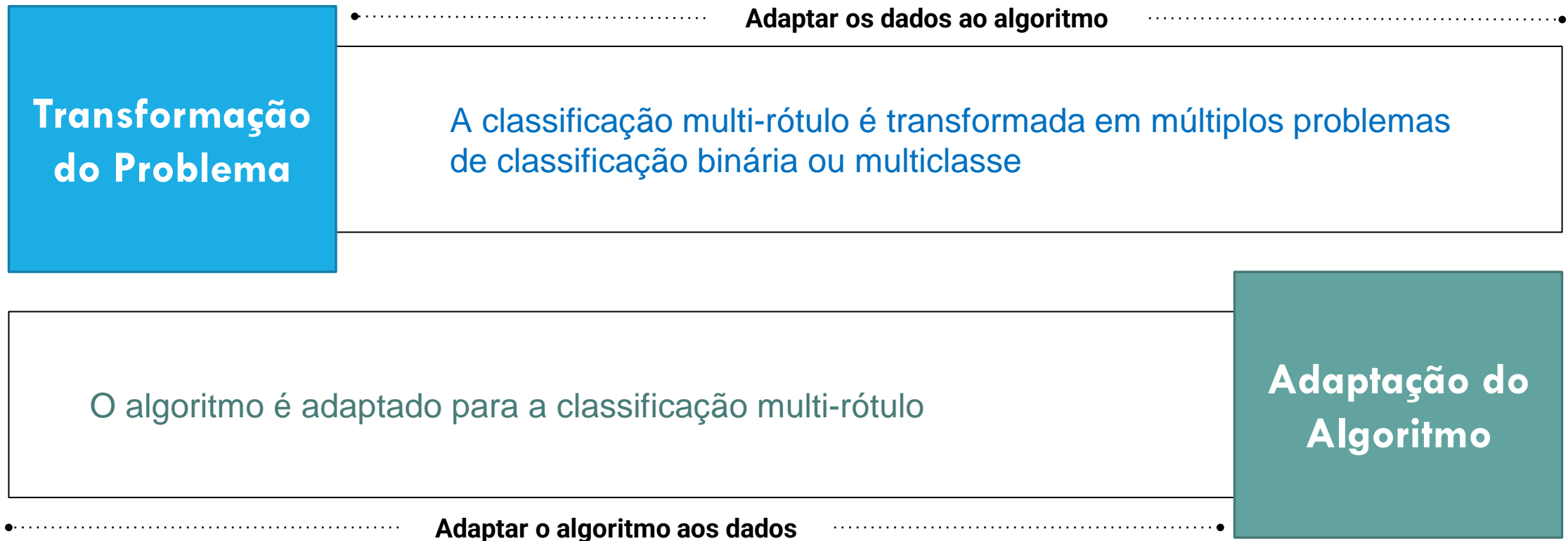
Refere-se aos problemas que permitem que seja atribuído múltiplas classes a uma única entrada (i.e., não mutuamente exclusivas)

A saída do sistema pode ser 1 ou mais classes

Baseado em duas abordagens básicas: transformação do problema (um classificador multiclasse para cada classe de rótulo)



CLASSIFICAÇÃO MULTI-RÓTULO: ABORDAGENS



Diversos métodos utilizam ambas as abordagens

CLASSIFICAÇÃO MULTI-RÓTULO



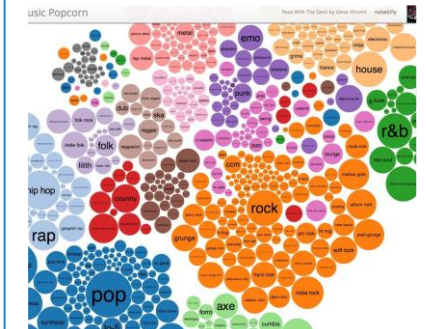
Rec de Gênero de Filmes



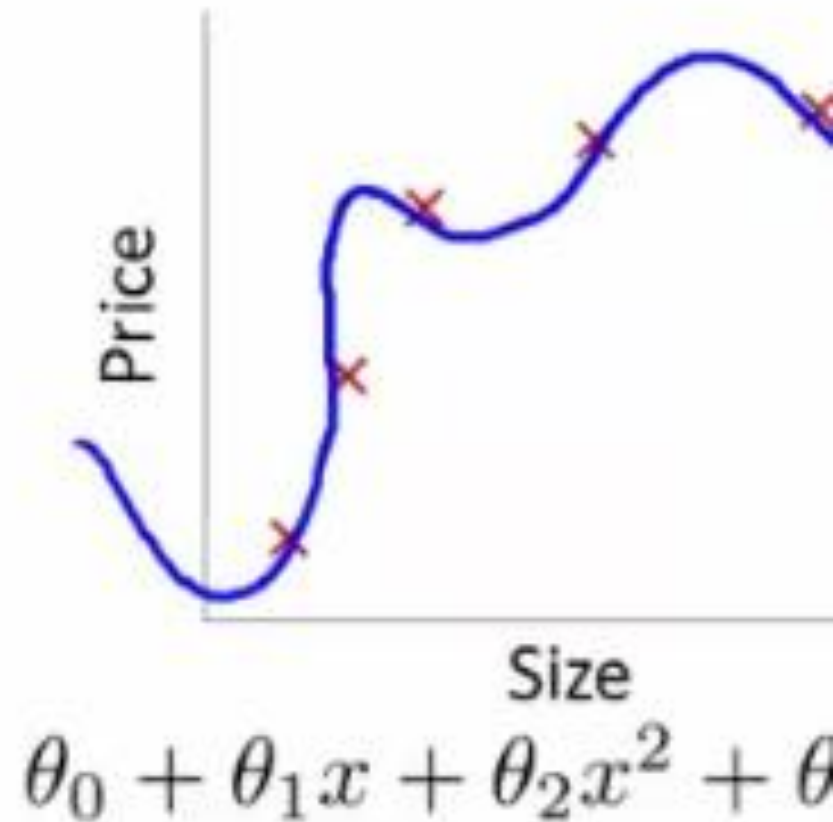
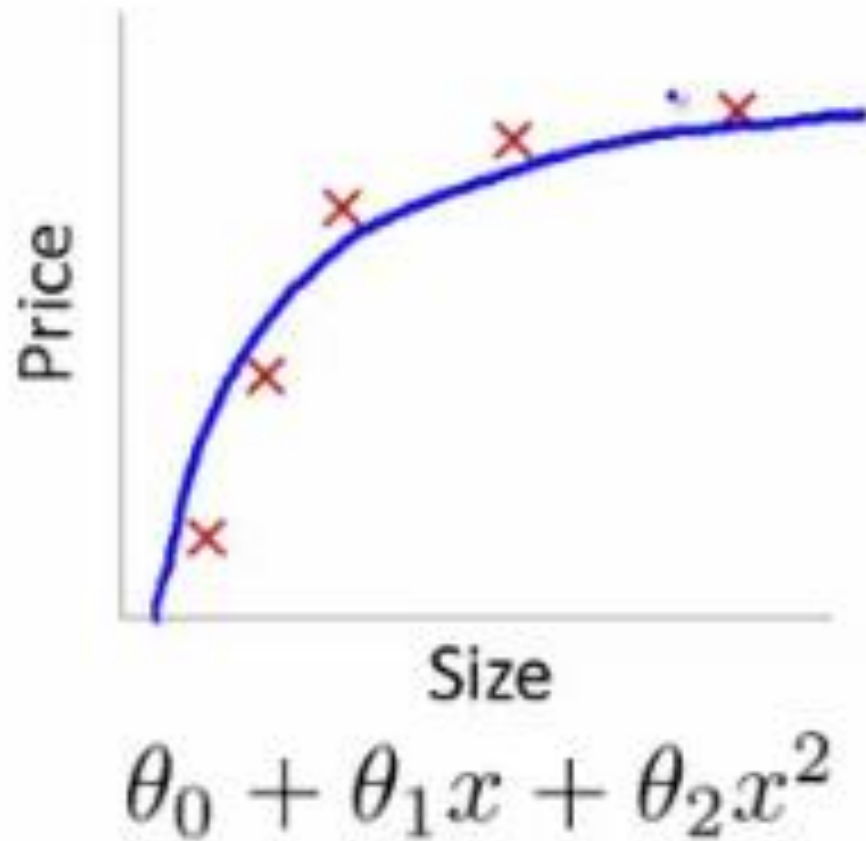
Predição de Risco de Doenças



Classificação de Documentos



Rec Gênero Musical



REGRESSÃO

Classes de Problemas
Supervisionado

REGRESSÃO

Processo de prever um valor numérico e contínuo para uma determinada amostra

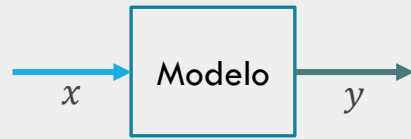


Geralmente utilizado para problemas do tipo previsão (determinar a cotação do dólar na próxima hora, quantas peças serão produzidas com falhas, quanto dias faltam até a próxima falha de um componente, máquina ou processo)

REGRESSÃO: TIPOS

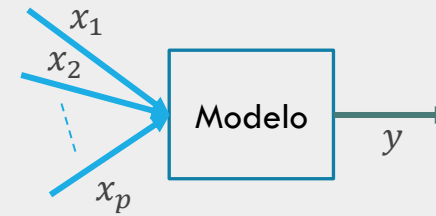
Unidimensional

Uma variável dependente y (contínua) é predita utilizando uma variável dependente x .



Multidimensional

Uma variável dependente y (contínua) é predita utilizando duas ou mais variáveis dependentes, x_1, x_2, \dots, x_p



O modelo captura o relacionamento entre a(s) variável(eis) de entrada e a de saída (não infere causalidade)

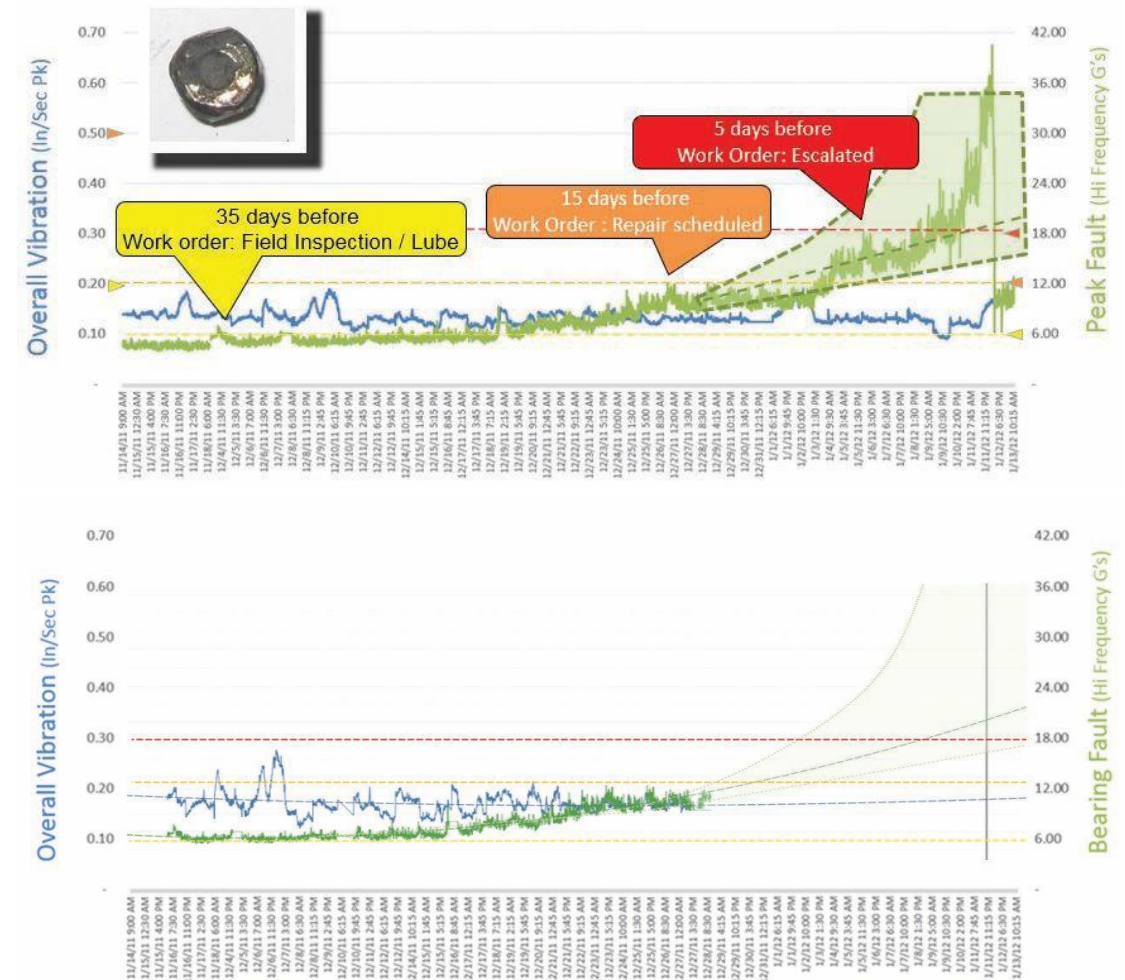
Possui um número limitado de métodos de regressão

REGRESSÃO: ANÁLISE PREDITIVA

Bombas, ventiladores e motores apresentam problemas de lubrificação, criação de cavidades, rolamentos, desnível, fixação

Medidas a partir de um acelerômetro permitem capturar informações sobre a vibração (2-1KHz e 1K-30KHz)

Algoritmos de regressão são aplicados para detectar problemas antes que aconteçam



<https://www.datascience.com/blog/predictive-analytics-in-industrial-iot>



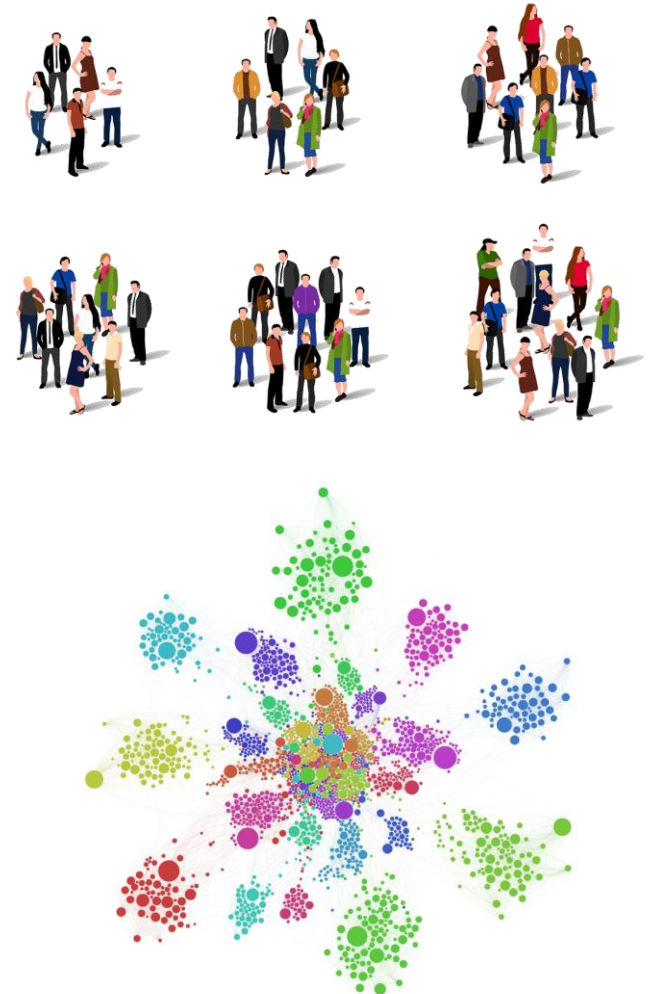
AGRUPAMENTO

Classes de Problemas
Não Supervisionado

AGRUPAMENTO (CLUSTERING)

Processo de arranjar os dados em grupos distintos (clusters), onde os membros de um grupo compartilham semelhanças entre si, mas não entre dados de grupos diferentes

Utilizada também para visualização de dados e pré-processamento para outros algoritmos de aprendizagem



AGRUPAMENTO: DESAFIOS

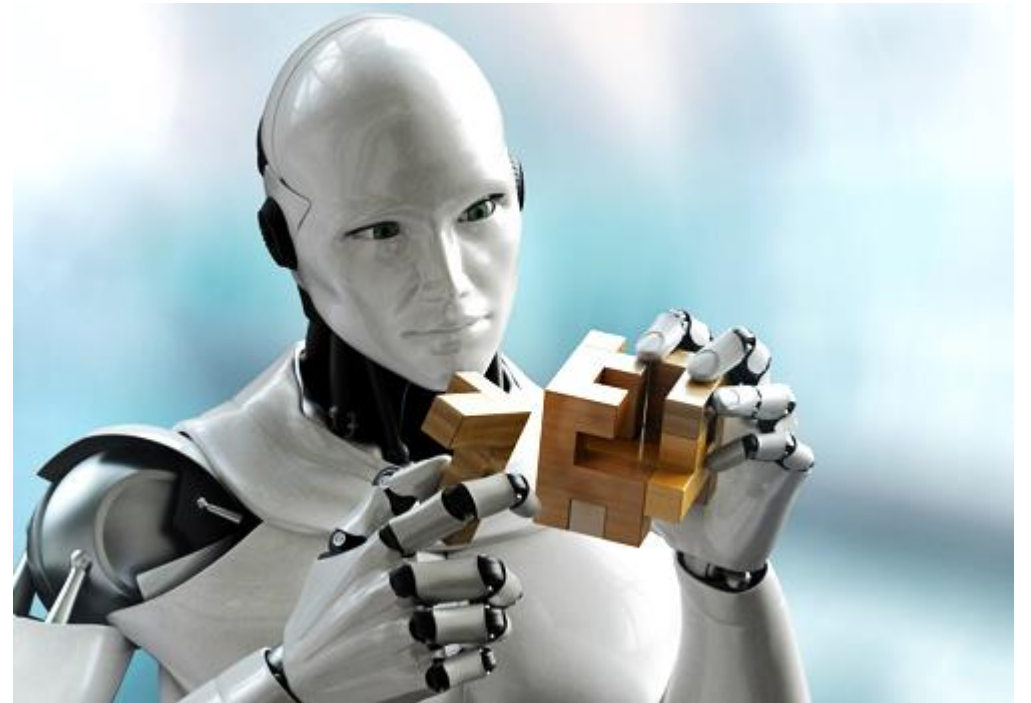
Quais medidas utilizar para avaliar a similaridades ou diferenças entre os dados?

Quantos grupos produzir?

Como avaliar o resultado?

Trabalhar com grandes quantidades dados (escalabilidade, velocidade, complexidade, ...)

Não há garantia de que existam grupos nos dados (mas ele vai achar algo...)

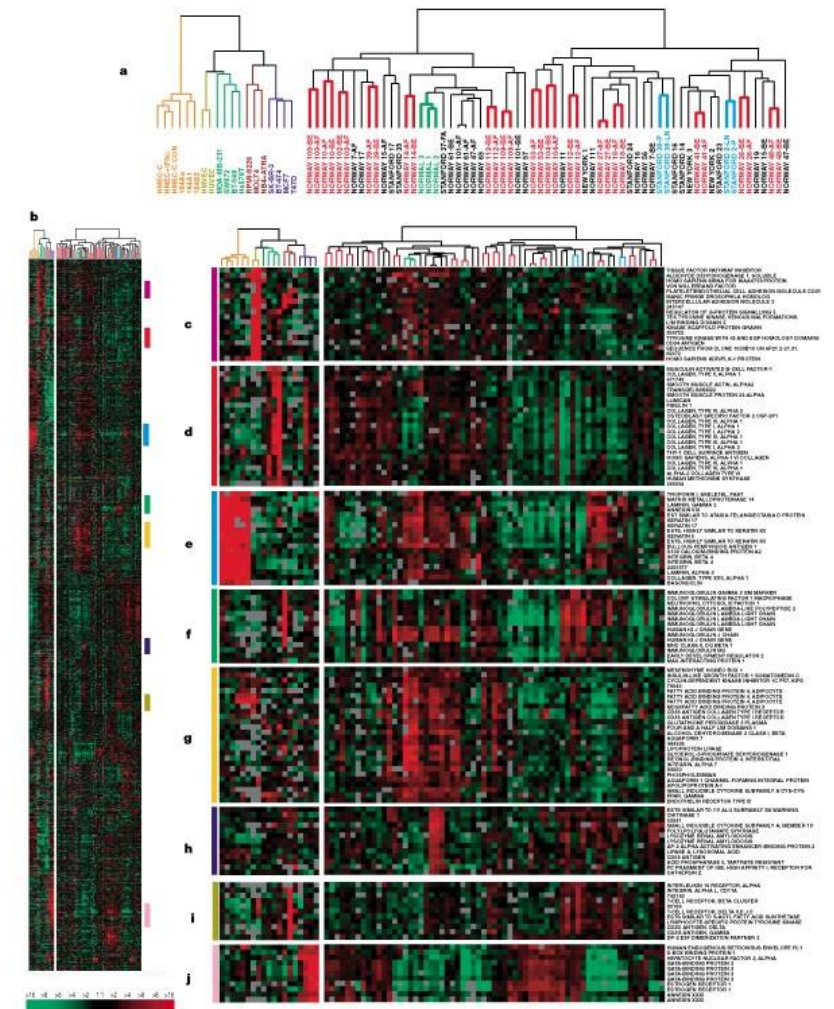


AGRUPAMENTO: APLICAÇÕES

Agrupamento tem sido extensivamente utilizado em estudos genéticos (visualização de dados ou detecção de subpopulações)

Estudos sobre câncer utilizam agrupamentos para detectar subtipos de tumores baseado nos padrões de expressão gênica*

* processo através do qual a informação contida nos genes é convertida em moléculas que determinam as propriedades da célula, ou seja, a transferência de informação genética de DNA a proteína designa-se por expressão gênica

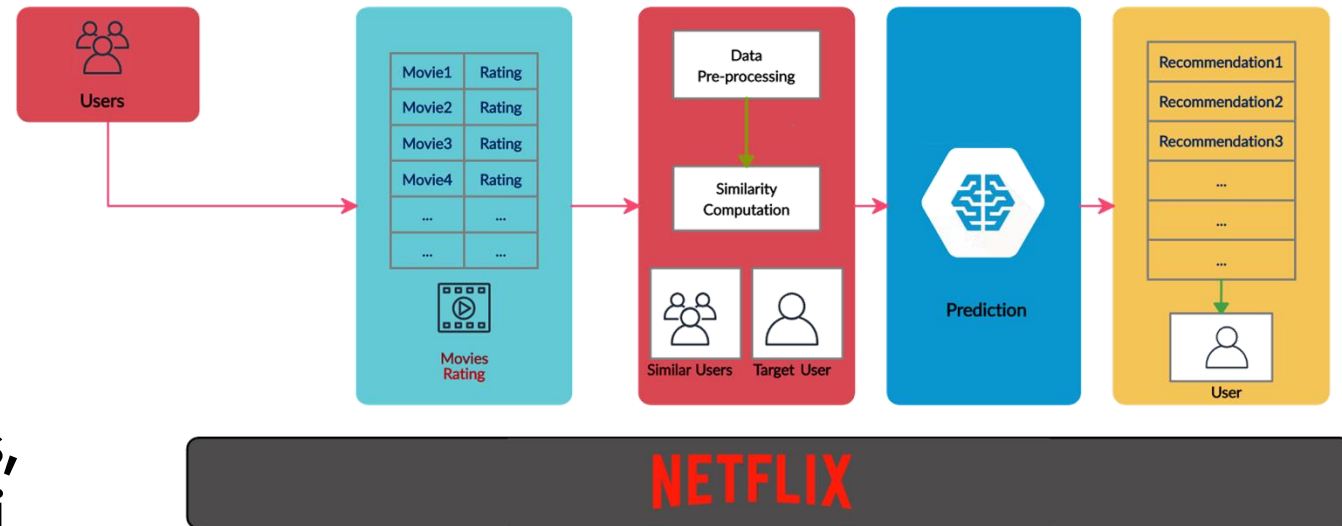


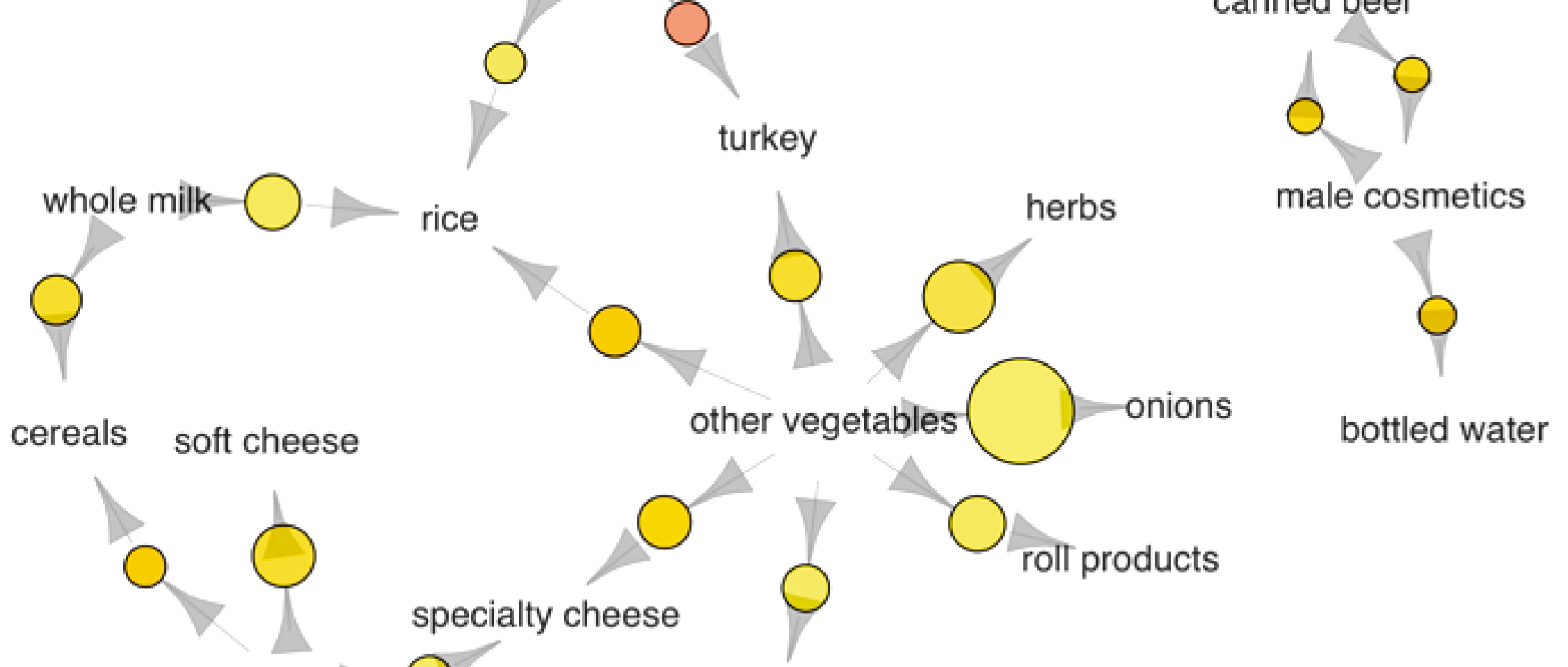
AGRUPAMENTO: APLICAÇÕES

O **sistema de recomendação** do Netflix utiliza agrupamentos para casar e buscar hábitos similares de usuário para sugerir filmes/séries que compartilham características com os filmes que foram bem avaliados por assinantes

O sistema combina classificação e agrupamento

São diversas fontes de informações: (avaliações, localização dos usuários, duração e hora que o filme/série foi assistida, dia da semana, todos os metadados do filme (incluindo as críticas), informações do registro e da mídia social do assinante, ...)





ASSOCIAÇÃO

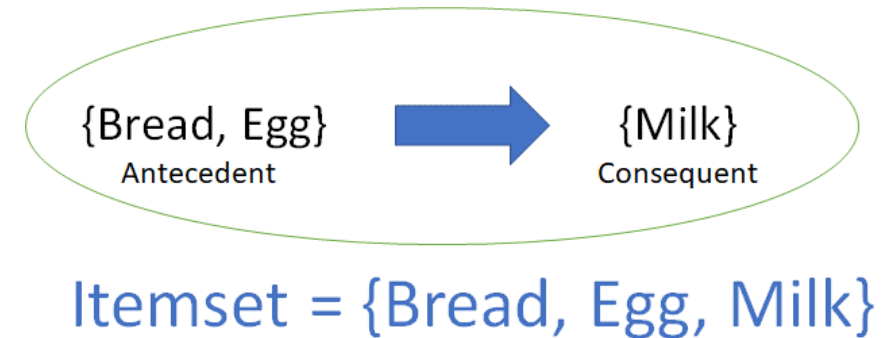
Classes de Problemas
Não Supervisionado

ASSOCIAÇÃO

Processo de descobrir padrões (regras) nos dados que permite descrever parte dos dados

A aprendizagem de associação busca relações importantes entre dados (diferente do agrupamento que busca similaridades)

Verifica a dependência entre itens de conjunto de dados que tenham determinada relevância (confiança)



ASSOCIAÇÃO: APLICAÇÕES

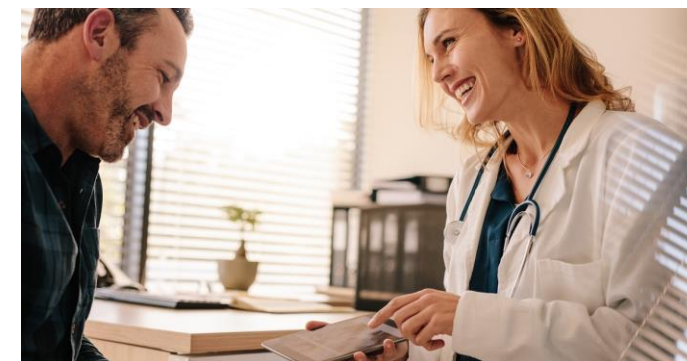
Análise do Carrinho de Compras

- Permite analisar o relacionamento entre produtos que nem sempre apresentam uma relação óbvia
- Pode auxiliar na disposição dos produtos em um mercado, o que usar em uma campanha, projetar um catálogo ou sugerir produtos



Diagnóstico Médico

- Regras de associação podem auxiliar os médicos no diagnóstico médico
- O diagnóstico pode ter o seu erro reduzido com base na probabilidade da ocorrência de uma doença conforme os diversos fatores e sintomas



RESUMINDO

A forma da aprendizagem é aplicada de acordo com a disponibilidade dos dados e o tipo de problema a ser resolvido

• Aprendizagem Supervisionada

- Demanda dados para realizar o aprendizado
- Para cada dado utilizado no aprendizado, é necessário incluir a saída esperada
- Para problemas de **Classificação**, a **saída** esperada é um **valor categórico** (conhecido também como rótulo)
- Para problemas de **Regressão**, a **saída** esperada é um **valor numérico e contínuo**

• Aprendizagem Não-Supervisionada

- Demanda dados para realizar o aprendizado
- Não possui a saída esperado, pois deseja-se encontrar padrões e estruturas no conjunto de dados

RESUMINDO

Os algoritmos de aprendizado são aplicáveis conforme a classe de problema

- **Aprendizagem Supervisionada**

Classificação: prever a classe (saída categórica) de uma determinada amostra. Podem ser classificados em unária, binária, multi-classe (conjunto aberto x fechado) e multi-rótulo

Regressão: previsão de um valor numérico e contínuo a partir de uma determinada amostra

- **Aprendizagem Não-Supervisionada**

Agrupamento: descobrir estruturas ou padrões nos dados

Associação: extração de regras que descrevem relações

RESUMINDO

As tarefas possuem nomes que dependem da tarefa ou do contexto onde é utilizado

- Detecção/Verificação de Locutor, Detecção/Reconhecimento Facial Locutor (classificação binária)
- Identificação de Locutor, Reconhecimento de Língua (classificação multi-classe com conjunto aberto ou fechado)
- Reconhecimento de Fala (classificação multi-classe)
- Sistema de Recomendação (agrupamento ou associação)
- Diagnóstico médico (classificação, agrupamento, ...)
- Detecção de anomalias / outliers / novidades (classificação unária)

A avaliação do resultado dependerá do tipo de problema e contexto a ser aplicado