

## Aprendizado de Máquina

Principais aspectos

André C. P. L. F. de Carvalho  
ICMC-USP



## Tópicos

- Aprendizado de Máquina
- O que faz?
- AM x computação convencional
- Aplicações
- Tarefas
- Abordagens

© André de Carvalho - ICMC/USP

2

## Aprendizado de Máquina (AM)

- Revolução industrial automatizou trabalho manual
- Revolução da informação automatizou trabalho mental
- Revolução de aprendizado de máquina automatiza a própria automação

André C. P. L. F. de Carvalho

3

## Riscos e benefícios

- Toda invenção, incluindo roda e fogo, pode ser beneficiar ou prejudicar a sociedade
  - Reduzir nossa exposição a situações de perigo
  - Reduzir a realização de tarefas repetitivas, monótonas e arriscadas
  - Melhorar a produtividade, permitindo fazer mais e melhor com menos esforço
  - Reduzir custos, permitindo que mais pessoas tenham acesso a um serviço ou produto
    - Saúde, educação, moradia, saneamento
  - Permitir mais tempo para família e lazer

© André de Carvalho - ICMC/USP

4

## Quando começou?

- Filósofo grego aristóteles criou a lógica moderna (384-322 AC)
  - Coleção de seis textos (organon: instrumento)
  - Reconhecia duas formas de argumentação
    - Dedução (raciocínio dedutivo)
      - A partir de um ou mais termos gerais chega-se a uma conclusão logicamente correta
      - Lógica top-down
    - Indução (raciocínio indutivo)
      - Generaliza a partir de conhecimentos particulares
      - Lógica bottom-up
      - **Aprendizado**



5

## Um dos marcos

- Isaac Newton publica *Principia* em 1687
  - Princípios Matemáticos da Filosofia Natural
    - Conjunto de 3 livros
      - 3 leis de movimento
        - Lei da inércia
        - Princípio fundamental da dinâmica
        - Lei da ação e reação



© André de Carvalho - ICMC/USP

6

## Um dos marcos

- Isaac Newton publica *Principia* em 1687
  - Principípios Matemáticos da Filosofia Natural
    - Conjunto de 3 livros
      - 3 leis de movimento
        - Lei da inércia
        - Princípio fundamental da dinâmica
        - Lei da ação e reação
      - 4 regras de indução
        - 3a regra: O que é verdade para tudo que vimos é verdade para tudo no Universo



© André de Carvalho - ICMC/USP

7

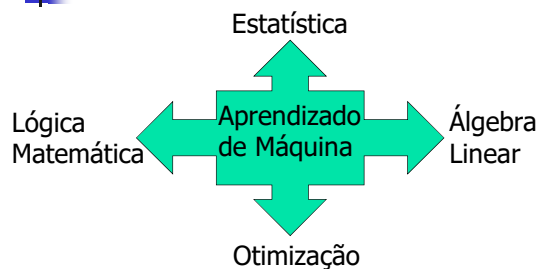
## O que faz? O que é?

- Investiga técnicas computacionais capazes de adquirir automaticamente
  - Novas habilidades, conhecimentos e formas de organizar o conhecimento existente
- Definições
  - Área de pesquisa que dá aos computadores a habilidade de aprender sem ser explicitamente programado (Samuel, 1959)
  - Técnicas capazes de melhorar seu desempenho em uma dada tarefa utilizando experiências prévias (Mitchell, 1997)

André C P L F de Carvalho

8

## É só computação?



© André de Carvalho - ICMC/USP

9

## AM vs computação tradicional

Escreve algoritmo detalhando como resolver um problema

Computação tradicional

Aprende a resolver um problema observando seus dados

Aprendizado de Máquina

André C P L F de Carvalho

10

## AM vs computação tradicional

Pessoas programam

Computação tradicional

Computadores programam

Aprendizado de Máquina

André C P L F de Carvalho

11

## AM vs computação tradicional

- Programação
  - Programa que funciona 90% das vezes é ruim
- Aprendizado de máquina
  - Modelo que acerta 90% das vezes pode ser o possível
    - E muito bom
    - Muitas vezes, é suficiente para ser útil

© André de Carvalho - ICMC/USP

12

## Aplicações de AM

- Programas baseados em AM têm sido bem sucedidos em:
  - Análise de redes sociais
  - Extração de conhecimento de dados biológicos
  - Detecção de fraudes
  - Diagnóstico médico
  - Biometria
  - Sistemas de recomendação

André C P L F de Carvalho

13

## Aplicações de AM

- AM esta presente em várias atividades do nosso dia-a-dia
- AM é utilizado para
  - Recomendar que mensagens mostrar em aplicativos de redes sociais
  - Filtrar spams de seus *emails*
  - Decidir que resultados (e anúncios) e em que ordem mostrar para uma busca na internet
  - Sugerir filmes ou livros que alguém vai gostar

© André de Carvalho - ICMC/USP

14

## Aplicações clássicas de AM

- Aprender a ler em voz alta
  - NETtalk (Terrence Sejnowski e Rosenberg, 1986)
- Aprender a reconhecer palavras faladas
  - SPHINX (Lee, 1989)
- Aprender a conduzir um automóvel
  - ALVINN (Pomerleau, 1989)
- Aprender a jogar gamão
  - TD-GAMMON (Tesauro 1992)

André C P L F de Carvalho

15

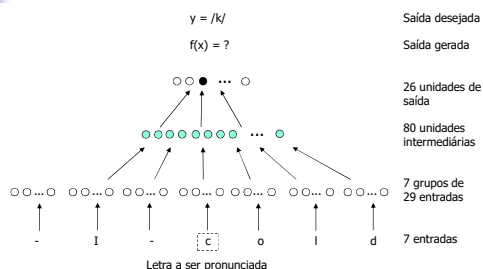
## NETtalk

- NETtalk: a parallel network that learns to read aloud*
  - Sistema automático para aprender a falar o que está escrito em um texto em inglês
  - Mapeia textos em fonemas
    - Sem usar processamento de língua natural nem regras da fonética
    - Usa rede neural com uma camada intermediária
      - Camada de entrada: 7 grupos com 29 entradas
      - Camada escondida: 80 neurônios
      - Camada de saída: 26 neurônios

André C P L F de Carvalho

16

## NETtalk



© André de Carvalho - ICMC/USP

17

## NETtalk

- Camada de entrada
  - Janela deslizante
  - 7 grupos para as sete posições do texto
    - 29 entradas = 26 letras + 3 marcas de pontuações/pausas
- Camada intermediária
  - Maioria dos neurônios reconhece mais de uma combinação de letras

André C P L F de Carvalho

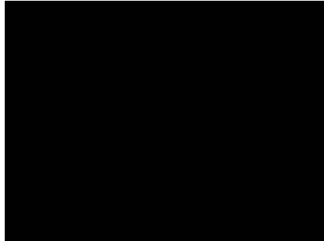
18

## NETtalk

- Camada de saída
  - Código do fonema para pronúncia da letra no centro da janela de entrada
    - Além de Silêncio, omissão de som (sílabas), pausa, final da leitura
  - Código gerado depende só das letras, não usa informação sobre fonemas vizinhos

André C P L F de Carvalho 19

## NETtalk



© André de Carvalho - ICMC/USP 20

## NETtalk

- Dados disponíveis no repositório da UCI
  - Lista de 20.008 palavras em inglês
    - Junto com a transcrição fonética de cada palavra
    - Cada exemplo tem 4 campos
      - Representação da letra
      - Representação do fonema
      - Sílabas tônicas e estrutura silábica
      - Valor inteiro indicando palavras estrangeiras e irregulares


© André de Carvalho - ICMC/USP 21

## NETtalk

- Desempenho
  - 95% para os dados de treinamento e 78% para os dados de teste
  - Comparado com Dectalk
    - Sistema especialista desenvolvido por linguistas e baseado em regras
    - Dectalk apresentou um desempenho melhor
      - Foi desenvolvido em uma década de análises por linguistas

© André de Carvalho - ICMC/USP 22

## ALVINN



Dean Pomerleau  
CMU

André C P L F de Carvalho 23

## ALVINN

- *Autonomous Land Vehicle In a Neural Network*
  - Sistema automático de navegação para automóveis baseado em redes neurais
    - Tese de doutorado da CMU
  - Comunicação por uma câmera montada no veículo
  - Dirigiu a 110 Km/h em uma rodovia pública americana em 1989
    - De costa a costa por 4500 Km (com exceção de 80 Km)

André C P L F de Carvalho 24

## ALVINN



© André de Carvalho - ICMC/USP

25

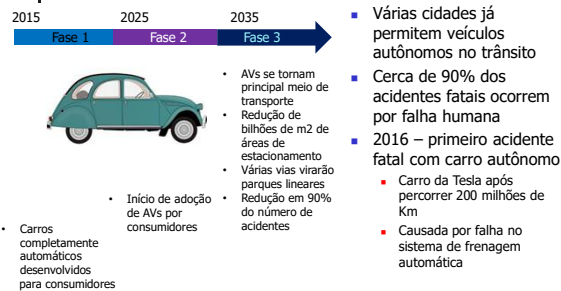
## Carros Tesla



© André de Carvalho - ICMC/USP

26

## Carros autônomos (AVs)



© André de Carvalho - ICMC/USP

27

## Conjuntos de dados

- Estruturados
  - Mais facilmente analisados por algoritmos de AM
  - Ex.: Planilhas e tabelas atributo-valor
- Não estruturados
  - Mais facilmente analisados por seres humanos
  - Ex.: Imagens, sequência de DNA, textos, páginas na web, emails, mensagens
    - Antes do uso de algoritmos de AM, são convertidos para estruturados

André C P L F de Carvalho

28

## Dados não estruturados



Girafa



Zebra

André C P L F de Carvalho

29

## Dados estruturados

|     |     |         |        |
|-----|-----|---------|--------|
| 500 | 110 | Manchas | Girafa |
| 440 | 90  | Manchas | Girafa |
| 240 | 45  | Listas  | Zebra  |
| 520 | 115 | Manchas | Girafa |
| 260 | 50  | Listas  | Zebra  |
| 230 | 50  | Listas  | Zebra  |

André C P L F de Carvalho

30

## Conjuntos de dados

- Rotulados
  - Cada objeto recebe um rótulo
    - Classe
    - Valor real
- Não rotulados
  - Objetos não recebem rótulos
- Parcialmente rotulados
  - Alguns objetos recebem rótulos

André C P L F de Carvalho

31

## Dados estruturados rotulados

| Atributos de entrada (preditivos)    |        |              |         |               |
|--------------------------------------|--------|--------------|---------|---------------|
|                                      | Altura | Tamanho Rabo | Textura | Classe        |
| Exemplos<br>(objetos,<br>instâncias) | 500    | 110          | Manchas | Girafa        |
|                                      | 440    | 90           | Manchas | Girafa        |
|                                      | 240    | 45           | Listas  | Zebra         |
|                                      | 520    | 115          | Manchas | Girafa        |
|                                      | 260    | 50           | Listas  | Zebra         |
|                                      | 230    | 50           | Listas  | Zebra         |
|                                      |        |              |         | Atributo alvo |

André C P L F de Carvalho

32

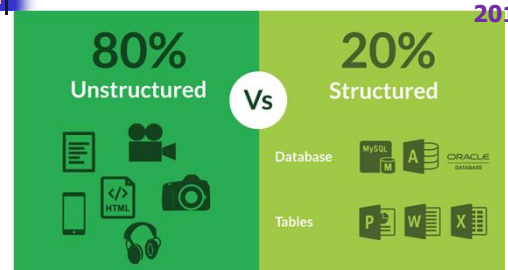
## Dados estruturados não rotulados

| Atributos de entrada (preditivos)    |        |              |         |
|--------------------------------------|--------|--------------|---------|
|                                      | Altura | Tamanho Rabo | Textura |
| Exemplos<br>(objetos,<br>instâncias) | 500    | 110          | Manchas |
|                                      | 440    | 90           | Manchas |
|                                      | 240    | 45           | Listas  |
|                                      | 520    | 115          | Manchas |
|                                      | 260    | 50           | Listas  |
|                                      | 230    | 50           | Listas  |

André C P L F de Carvalho

33

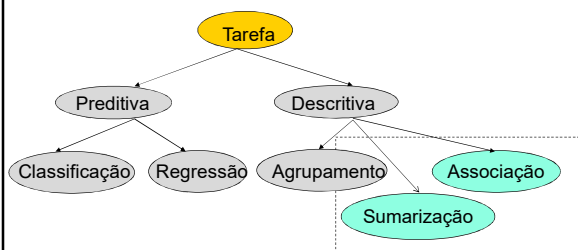
## Dados (não) estruturados


<https://upxacademy.com/beginners-guide-to-big-data/>

© André de Carvalho - ICMC/USP

34

## Tarefas de aprendizado



André C P L F de Carvalho

35

## Aprendizado pode ser

- Supervisionado
  - Tarefa preditiva (mais comum) ou descritiva
  - Modelo aprende tendo acesso a saída desejada
    - Fornece, para cada entrada, a saída desejada (correta)
- Não supervisionado
  - Tarefa descritiva (mais comum) ou preditiva
  - Aprendizado usa apenas atributos preditivos
- Semi-supervisionado (aprendizado ativo)
- Reforço

André C P L F de Carvalho

36

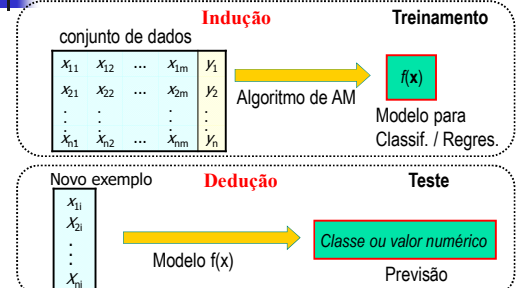
## Algoritmos para tarefas preditivas

- Induzem modelos (funções) preditivas
  - Dados de treinamento rotulados
- Modelo pode ser aplicado a novos dados (predição)
  - Rotula dados de teste
- Principais tarefas:
  - Classificação
  - Regressão

André C P L F de Carvalho

37

## Algoritmos para tarefas preditivas



André C P L F de Carvalho

38

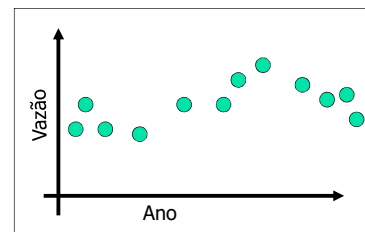
## Regressão

- Objetivo: aprender uma função que mapeia descrição de um exemplo em um valor real
- Exemplos:
  - Prever valor de mercado de um imóvel
  - Prever o lucro de um empréstimo bancário
  - Prever tempo de internação de paciente

André C P L F de Carvalho

39

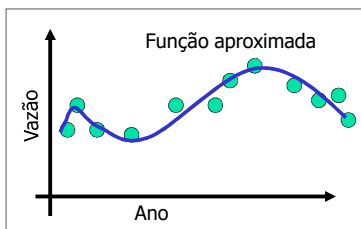
## Regressão



André C P L F de Carvalho

40

## Regressão

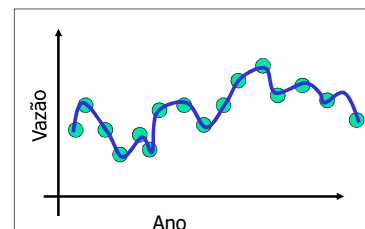


André C P L F de Carvalho

41

## Regressão

- Overfitting

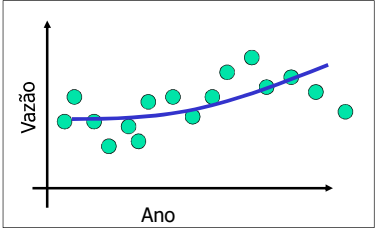


© André de Carvalho - ICMC/USP

42

## Regressão

- Underfitting



© André de Carvalho - ICMC/USP

## Regressão

- Técnicas
  - Árvores de Regressão
  - Redes Neurais Artificiais
  - Máquinas de Vetores de Suporte
  - Regressão Linear

André C P L F de Carvalho

## Classificação

- Objetivo: aprender uma função que associa descrição de um exemplo a uma classe
- Exemplos:
  - Definir a função de uma proteína
  - Distinguir emails entre spam e ham
  - Definir se um paciente tem ou não uma doença

André C P L F de Carvalho

## Classificação

- Posto médico A
  - Tem um histórico de vários atendimentos e diagnósticos
  - João, ao sentir alguns sintomas, vai ao posto para uma consulta médica
  - O único médico, faltou
    - Mas uma enfermeira pode anotar os sintomas
  - É possível fazer um pré-diagnóstico?

André C P L F de Carvalho

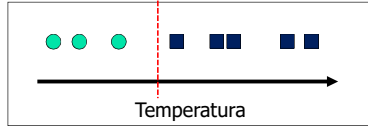
## Classificação

- Sintomas coletados pela enfermeira:
  - Temperatura

André C P L F de Carvalho

## Classificação

- Forma mais simples

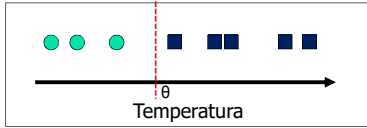


André C P L F de Carvalho



## Classificação

- Forma mais simples



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

- Se temperatura  $> \theta$   
Então doente  
Senão saudável

André C P L F de Carvalho 49

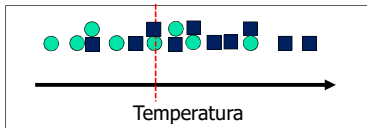
## Classificação

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

André C P L F de Carvalho 50

## Classificação

- Problema pode não ser tão simples



- Alternativa: considerar outros sintomas para o diagnóstico

André C P L F de Carvalho 51

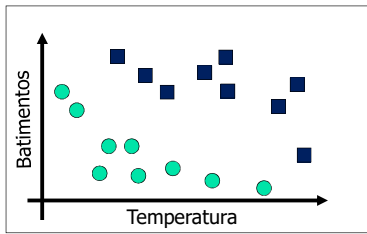
## Classificação

- Sintomas coletados pela enfermeira:
  - Batimentos cardíacos
  - Temperatura

André C P L F de Carvalho 52

## Classificação

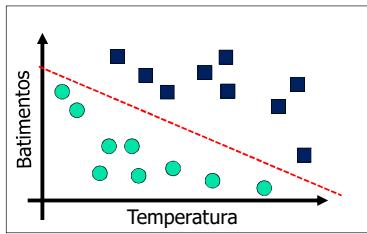
- Incluir número de batimentos



André C P L F de Carvalho 53

## Classificação

- Função linear permite diagnóstico



Nova função:  
Se  $a \cdot t + b > 0$   
Então doente  
Senão saudável

André C P L F de Carvalho 54

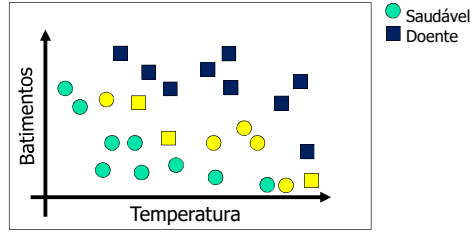
## Classificação

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

André C P L F de Carvalho 55

## Classificação

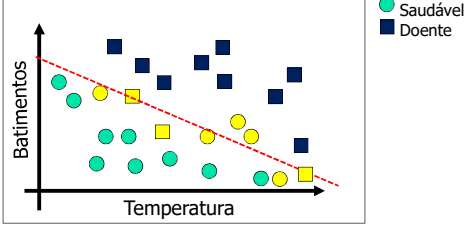
- Supor inclusão de outros pacientes



André C P L F de Carvalho 56

## Classificação

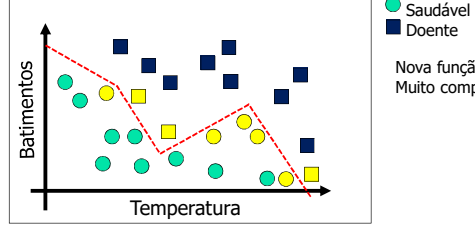
- Supor inclusão de outros pacientes



André C P L F de Carvalho 57

## Classificação

- Supor inclusão de outros pacientes

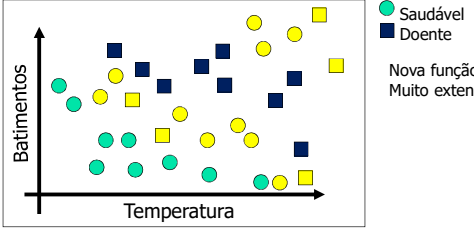


Nova função:  
Muito complexa

André C P L F de Carvalho 58

## Classificação

- Supor inclusão de mais pacientes

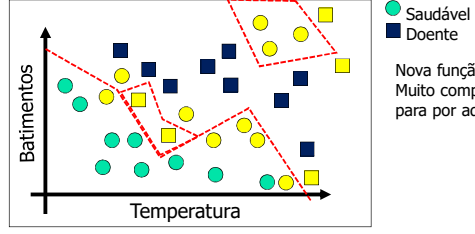


Nova função:  
Muito extensa

André C P L F de Carvalho 59

## Classificação

- Supor inclusão de outros pacientes



Nova função:  
Muito complexa para por aqui

André C P L F de Carvalho 60

## Algoritmos para tarefas descritivas

- Geram modelos em um processo de treinamento
  - Descrevem ou resumizam dados
  - Treinamento utiliza todos o conjunto de dados
  - Ex.: Agrupamento de dados
- Alguns algoritmos não utilizam treinamento
  - Ex.: Regras de associação e sumarização

© André de Carvalho - ICMC/USP

61

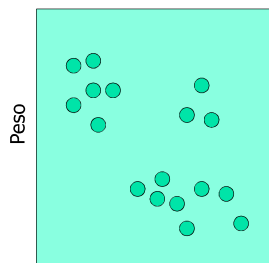
## Agrupamento (Clustering)

- Objetivo: organizar exemplos não rotulados em grupos (clusters)
  - De acordo com uma medida de similaridade ou correlação entre eles
  - Aprendizado não supervisionado
- Não existe conhecimento anterior sobre:
  - Número de grupos (várias vezes)
  - Significado dos grupos

© André de Carvalho - ICMC/USP

62

## Agrupamento

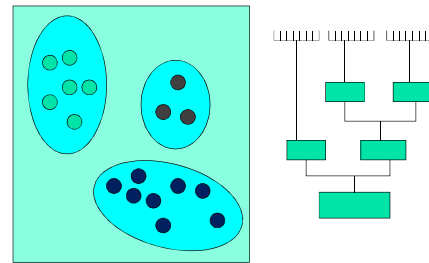


Exame 1

© André de Carvalho - ICMC/USP

63

## Agrupamento



© André de Carvalho - ICMC/USP

64

## Agrupamento

- Técnicas
  - Redes Neurais SOM
  - K-médias
  - FCM
  - DBSCAN
  - Single-Link

© André de Carvalho - ICMC/USP

65

## Sumarização

- Objetivo: encontrar descrição simples e compacta para um conjunto de dados
- Frequentemente utilizada para:
  - Exploração interativa de dados
  - Geração automática de relatórios
- Exemplo:
  - Definir perfis de compras feitas nos finais de semana em um supermercado

André Ponce de Leon F. de Carvalho

66

## Sumarização

| Nome  | Idade | Sexo | Altura | Tem filhos |
|-------|-------|------|--------|------------|
| João  | 32    | M    | 180    | S          |
| Maria | 30    | F    | -----  | N          |
| Pedro | 23    | M    | 160    | S          |
| José  | 45    | M    | 170    | S          |
| Sueli | 18    | F    | 175    | N          |

André Ponce de Leon F de Carvalho 67

## Sumarização

| Nome  | Idade | Sexo | Altura | Tem filhos |
|-------|-------|------|--------|------------|
| João  | 32    | M    | 180    | S          |
| Maria | 30    | F    | -----  | N          |
| Pedro | 23    | M    | 160    | S          |
| José  | 45    | M    | 170    | S          |
| Sueli | 18    | F    | 175    | N          |

Idade média: 29.6  
 Mediana da idade: 30  
 Sexo mais frequente: M  
 Maior altura: 180

André Ponce de Leon F de Carvalho 68

## Regras de Associação

- Objetivo: dado um conjunto de itens e uma base de dados de transações
  - Encontrar um conjunto de regras de associação entre os itens
- Exemplo:
  - Procurar por itens que são frequentemente comprados juntos
  - Itens frequentes

André Ponce de Leon F de Carvalho 69

## Regras de Associação

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

André Ponce de Leon F de Carvalho 70

## Regras de Associação

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


40% dos clientes compram pão e queijo  
 75% dos clientes que compram queijo também compram massa

André Ponce de Leon F de Carvalho 71

## Tópicos


- Aprendizado de Máquina
- O que faz?
- AM x computação convencional
- Aplicações
- Tarefas
- Abordagens

© André de Carvalho - ICMC/USP 72



## Perguntas

---



© André de Carvalho - ICMC/USP

73