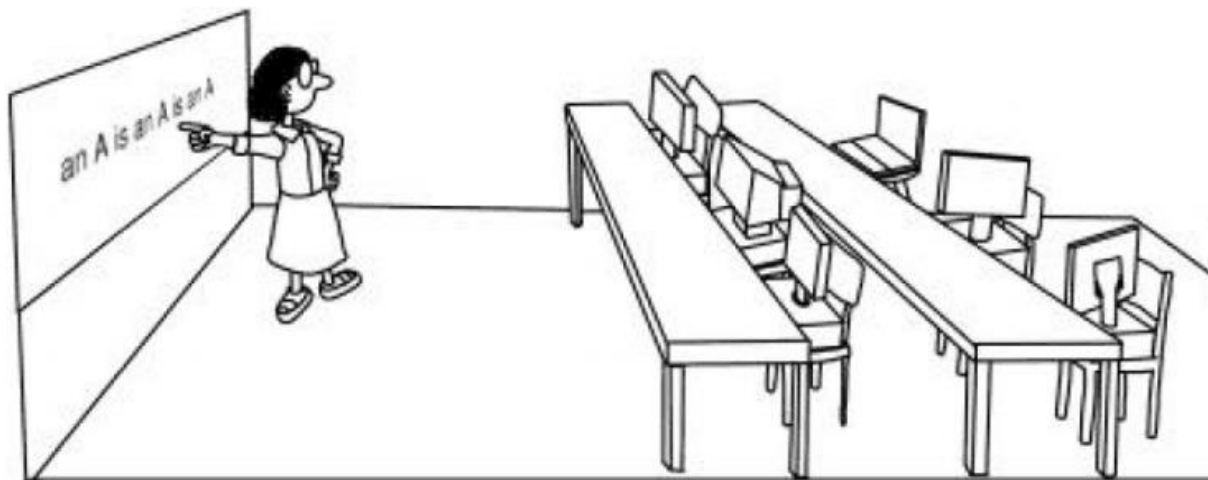


Curso Aprendizado de Máquina

Vamos falar um pouco sobre
tipos de aprendizado

O que é Aprendizagem de Máquina?



O que é Aprendizagem de Máquina?

- Aprender significa “**mudar para fazer melhor**” (de acordo com um dado critério) quando uma situação similar acontecer.
- Aprendizagem, **não é memorizar**. Qualquer computador pode memorizar, a dificuldade está em **generalizar** um comportamento para uma nova situação.

Importância do Aprendizado

- **Por que é importante para um agente aprender?**

- Os programadores não podem antecipar todas as situações que o agente pode encontrar.

Exemplo: Um robô programado para andar em um único labirinto pode não saber andar em outros.

- Os programadores não podem antecipar todas as mudanças que podem acontecer com o passar do tempo.

Exemplo: Agente programado para prever as melhores opção de bolsa para investir precisa se adapta quando o ambiente muda.

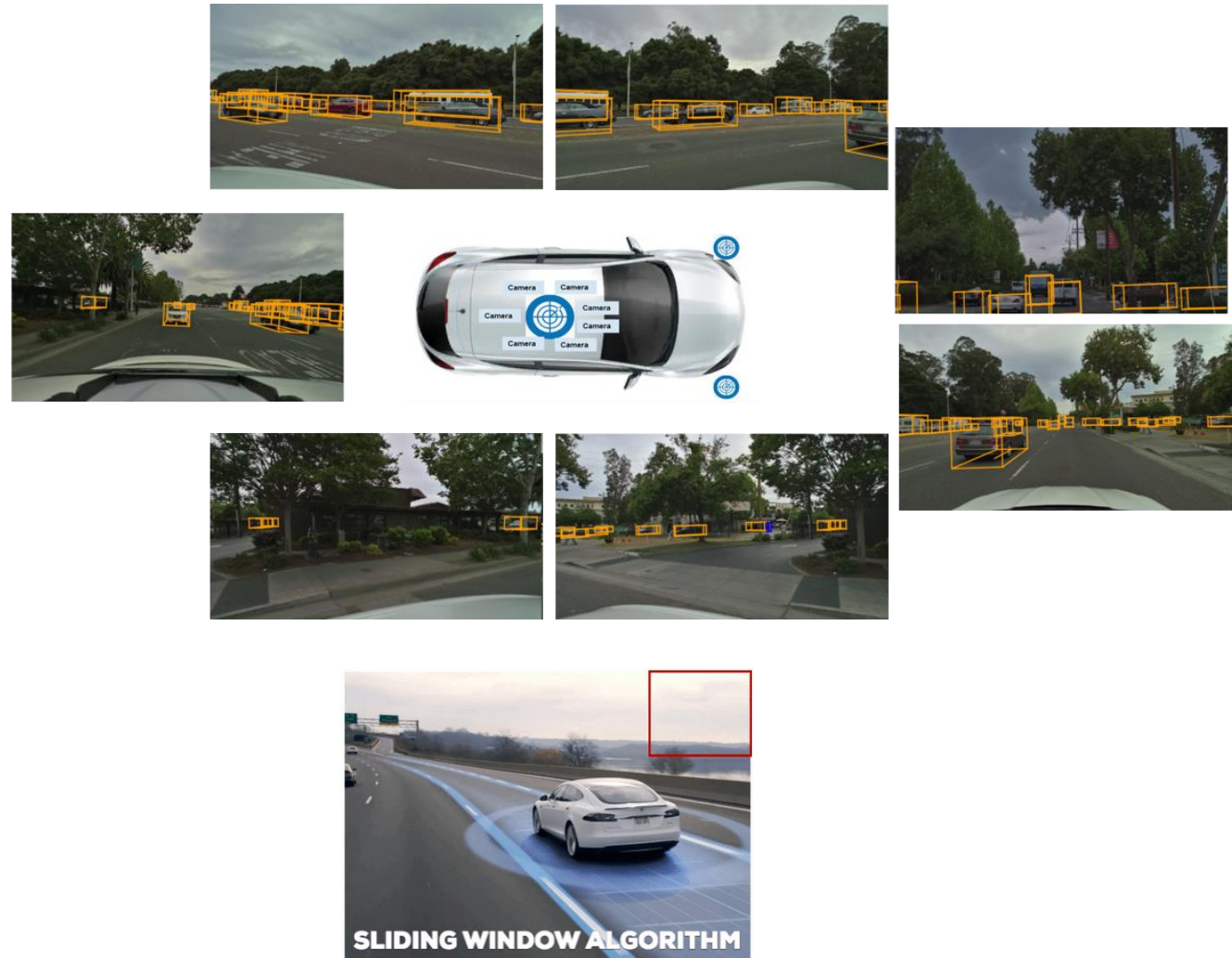
- Os programadores nem sempre sabem encontrar a solução dos problemas diretamente.

Exemplo: Programar um sistema para reconhecer faces não é algo trivial.

Como Aprender Algo?

- **Exemplos:**

- Considerando um agente treinando para se tornar um motorista de táxi. Toda vez que o instrutor gritar "freio!" o agente pode aprender uma condição de quando ele deve frear.
- Ao ver várias imagens que contêm ônibus, o agente pode aprender a reconhecê-los.
- Ao tentar ações e observar os resultados. Por exemplo, ao frear forte em uma estrada molhada pode aprender que isso não tem um efeito bom.



Formas de Aprendizado

- Aprendizado Supervisionado
- Aprendizado Não Supervisionado
- Aprendizado Por Reforço

Aprendizado Supervisionado

- Observa-se alguns pares de **exemplos de entrada e saída**, de forma a aprender uma **função que mapeia a entrada para a saída**.
- Damos ao sistema a “**resposta correta**” durante o processo de treinamento.
- É eficiente pois o sistema pode trabalhar diretamente com informações corretas.
- Útil para classificação, regressão, estimação de probabilidade condicional (qual é a probabilidade de um cliente com um determinado perfil comprar um determinado produto?)

Aprendizado Supervisionado

- **Exemplo:**

- Considerando um agente treinando para ser se tornar um motorista de táxi. Toda vez que o instrutor gritar "freio!" o agente pode aprender uma condição de quando ele deve frear.
- A entrada é formada pelos dados percebidos pelo agente através de sensores. A saída é dada pelo instrutor que diz quando se deve frear, virar a direita, virar a esquerda, etc.

Aprendizado Não Supervisionado

- O agente **reconhece padrões nos dados de entrada**, mesmo sem **nenhum feedback de saída**.
- Por exemplo, um agente aprendendo a dirigir pode gradualmente desenvolver um conceito de dias de bom tráfego e dias de tráfego congestionado mesmo sem nunca ter recebido exemplos rotulados por um professor.

Aprendizado Por Reforço

- O agente recebe uma série de reforços, **recompensas ou punições**.
- Por exemplo, a falta de uma gorjeta no final do percurso da ao agente taxista uma indicação de que ele fez algo errado.
- Cabe ao agente reconhecer qual das ações antes do reforço foram as maiores responsáveis por isso.
- Não damos a “resposta correta” para o sistema. O sistema faz uma hipótese e determina se essa hipótese foi boa ou ruim.

Fases da Aprendizagem

- **Treinamento**

- Apresenta-se exemplos ao sistema.
- O sistema “aprende” a partir dos **exemplos**.
- O sistema modifica gradualmente os seus parâmetros para que a saída se aproxime da saída desejada.

- **Utilização**

- Novos exemplos jamais visto são apresentados ao sistema.
- O sistema deve generalizar e reconhecê-los.

Algoritmos

- **Aprendizado Supervisionado**
 - Árvores de Decisão
 - KNN
 - SVM
 - Redes Neurais
- **Aprendizado Não Supervisionado**
 - Clusterização Sequencial
 - Clusterização Hierárquica
 - K-Means
- **Aprendizado Por Reforço**

Vamos lembrar nossa ultima
aula

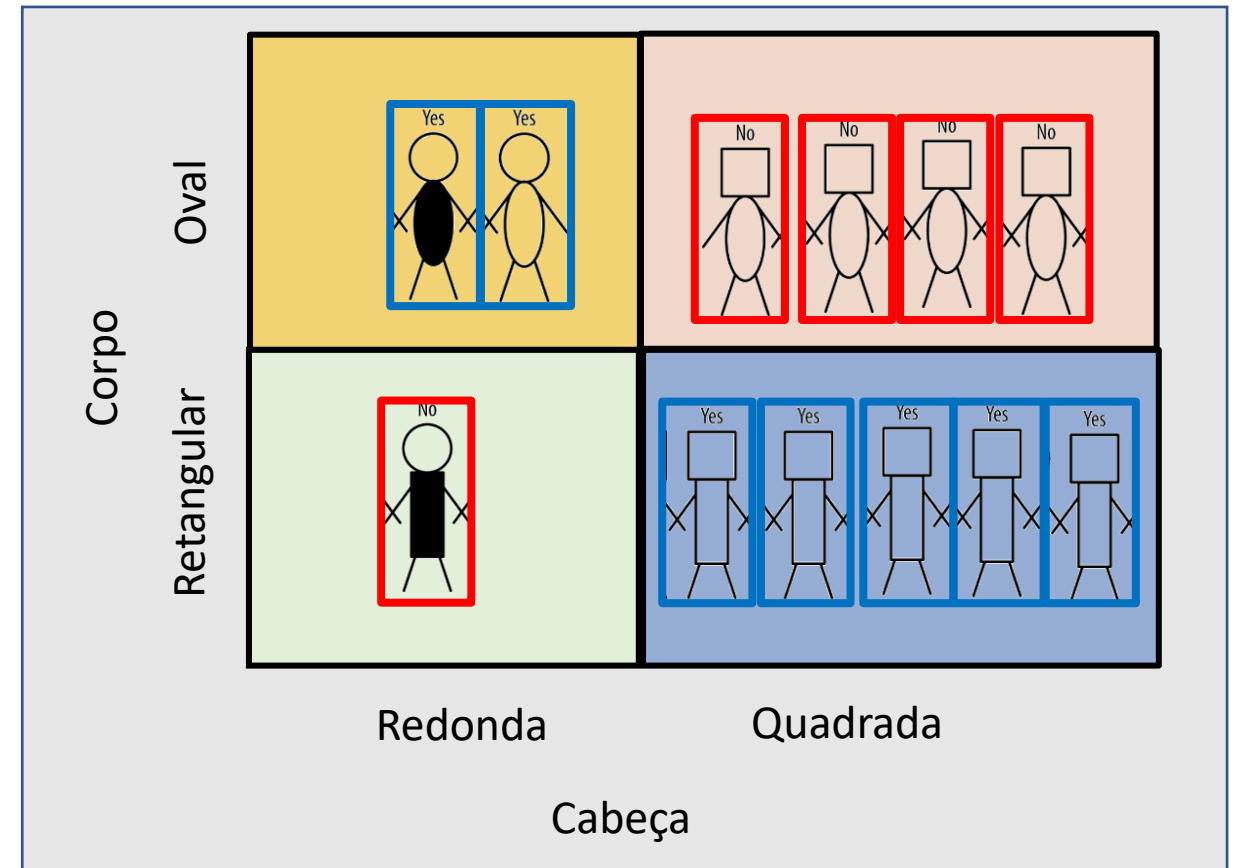
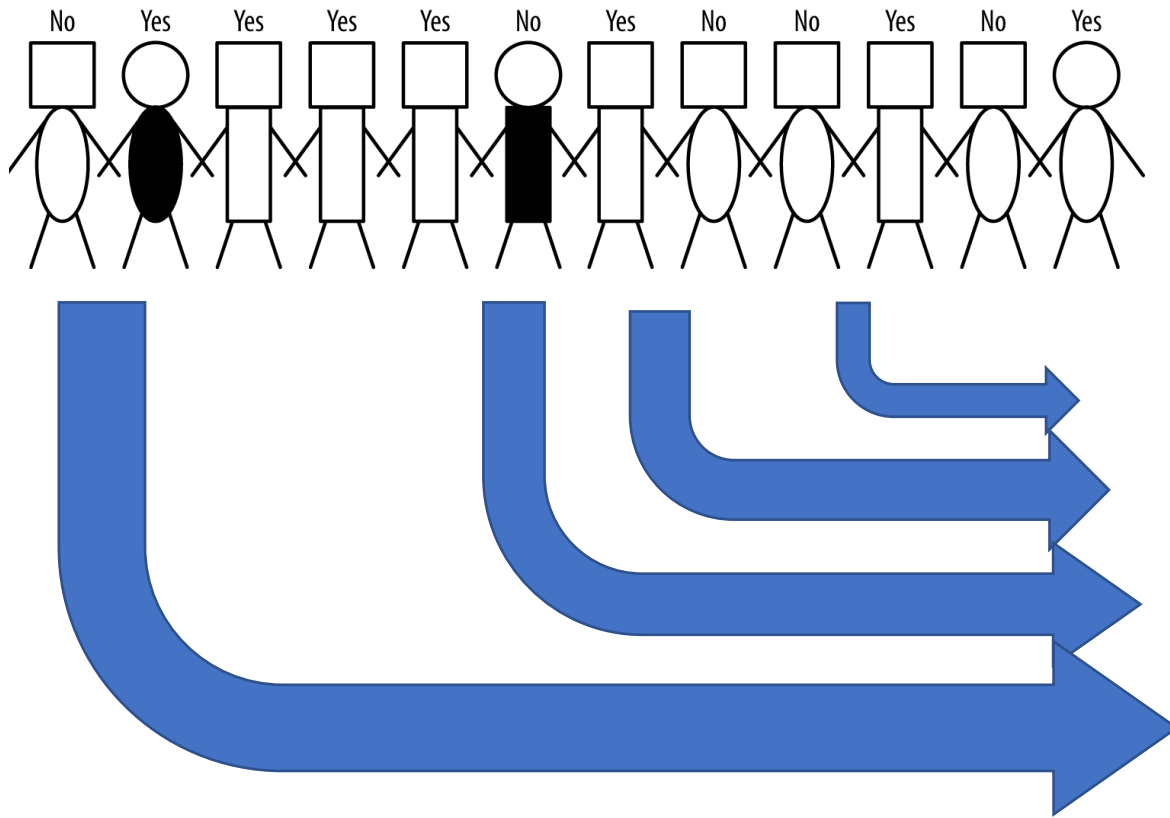
O conceito de informação

- No processo de discutir a segmentação supervisionada, apresentamos uma das idéias fundamentais é:
- Encontrar ou selecionar **variáveis importantes** ou **informativas** ou **"atributos"** das entidades descritas pelos **dados**.
- O que exatamente significa ser “informativo”?

O que exatamente significa ser “informativo”?

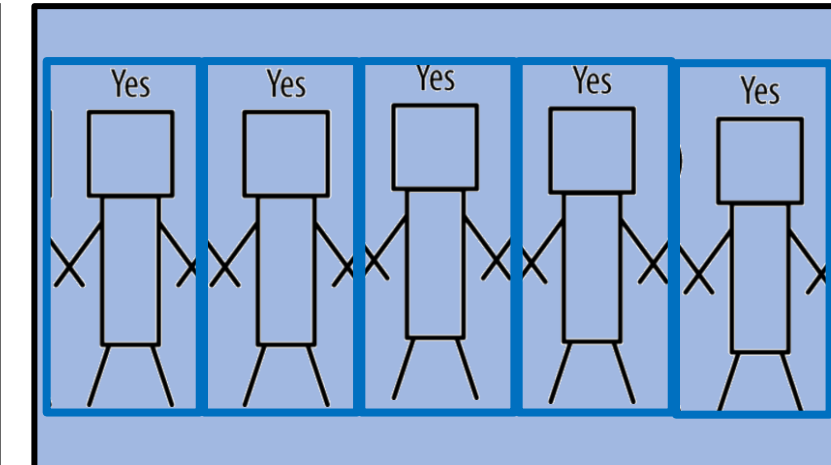
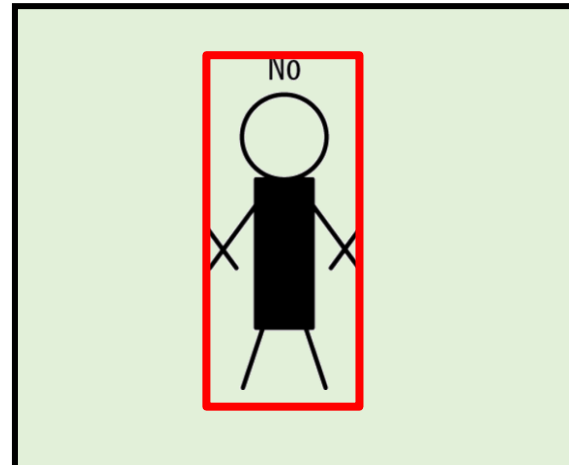
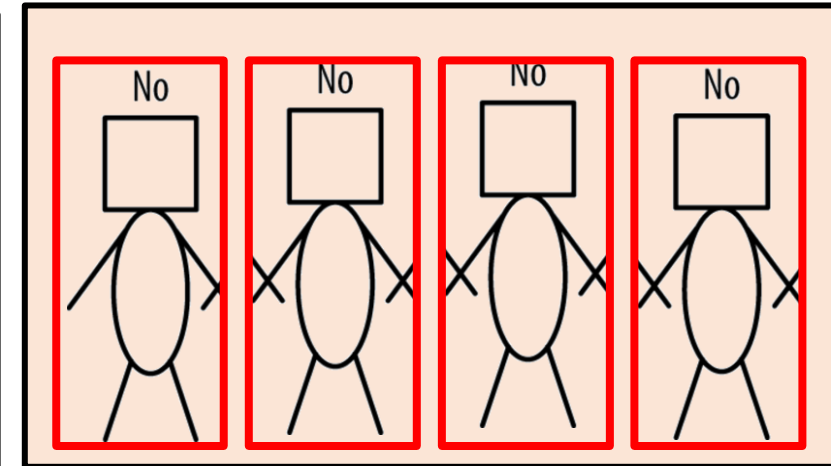
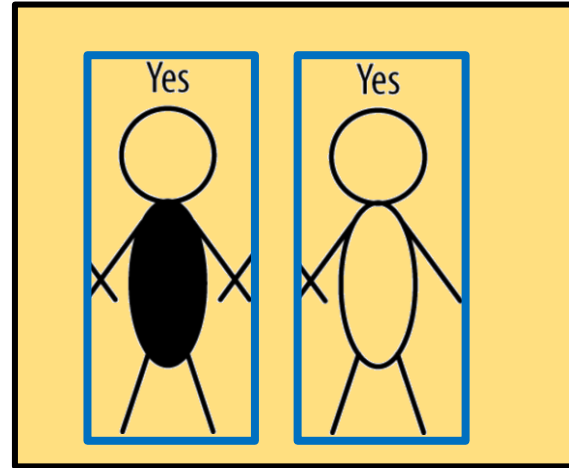
- Varia entre aplicativos, mas geralmente, a informação é uma quantidade que **reduz a incerteza** sobre algo.
- Portanto, se um velho pirata me fornece informações sobre onde está o tesouro dele, o que não significa que eu saiba com certeza onde está, isso significa apenas que minha incerteza sobre o local está escondida é reduzida.
- Quanto melhor a informação, mais minha incerteza é reduzida.

Separando em Grupos



Separando em Grupos

- Na aula anterior nós usamos um algoritmo baseado em entropia para separar as instâncias em grupos
- Esse processo foi baseado em organizar melhor a informação com base no atributo alvo crédito
- Portanto fizemos aprendizado supervisionado



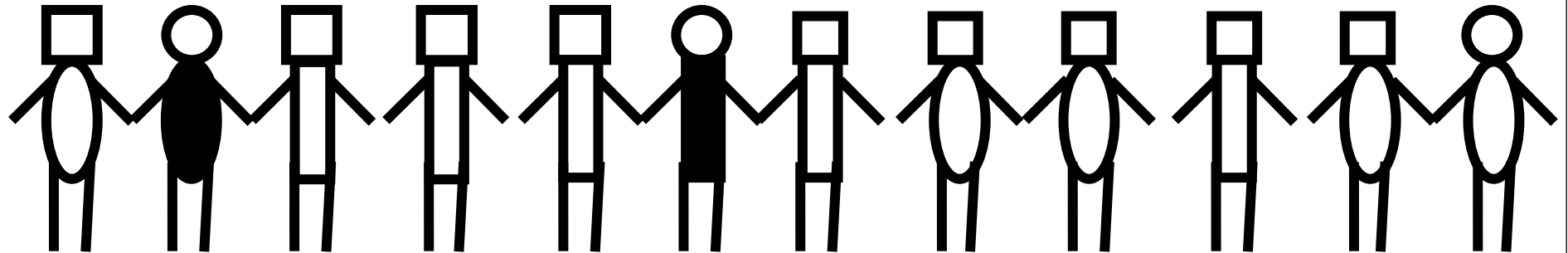
Usando a indução de árvores no nosso data set

- Elegemos um atributo alvo o Crédito, que nos ajudou a encontrar formas de organizar a informação em quatro grupos onde a entropia é menor
- Identificamos algumas regras que podem ser usadas para prever novos dados de crédito para novos clientes
- A única coisa que precisamos saber é o formato da cabeça e o formato do corpo e com isso podemos prever se o cliente irá ou não receber o crédito
- Essas regras são:
 - Se o Corpo Retangular e Cabeça Quadrada então SIM
 - Se o Corpo Retangular e Cabeça Redonda então NÃO
 - Se o Corpo Oval e Cabeça Quadrada então NÃO
 - Se o Corpo Oval e Cabeça Redonda então SIM

Vamos voltar ao nosso conjunto
de dados

Como está dividido nosso dataset

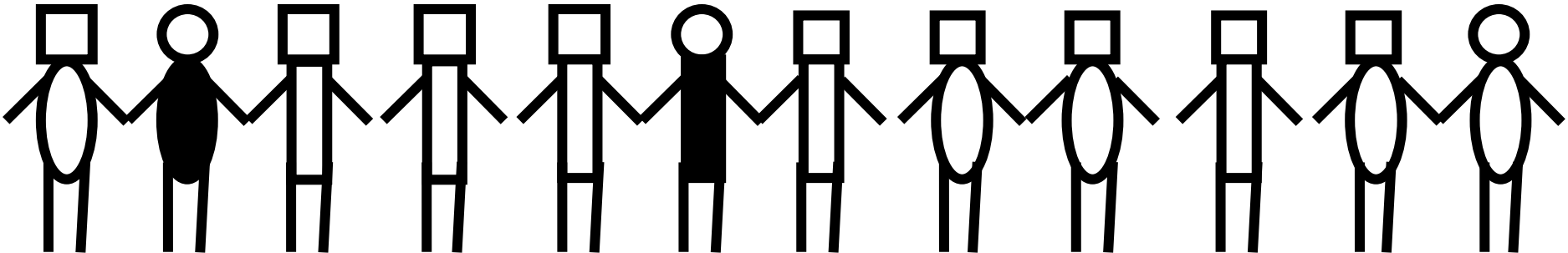
Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	Oval	Oval	Ret	Ret	Ret	Ret	Ret	Oval	Oval	Ret	Oval	Oval
Cabeça	Quad	Red	Quad	Quad	Quad	Red	Quad	Quad	Quad	Quad	Quad	Red
Cor do Corpo	Branco	Preto	Branco	Branco	Branco	Preto	Branco	Branco	Branco	Branco	Branco	Branco
Crédito	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Sim	Não	Sim



Vamos transformar nosso dataset – Fase 1

Corpo Retangular = 1 Corpo Oval = 0

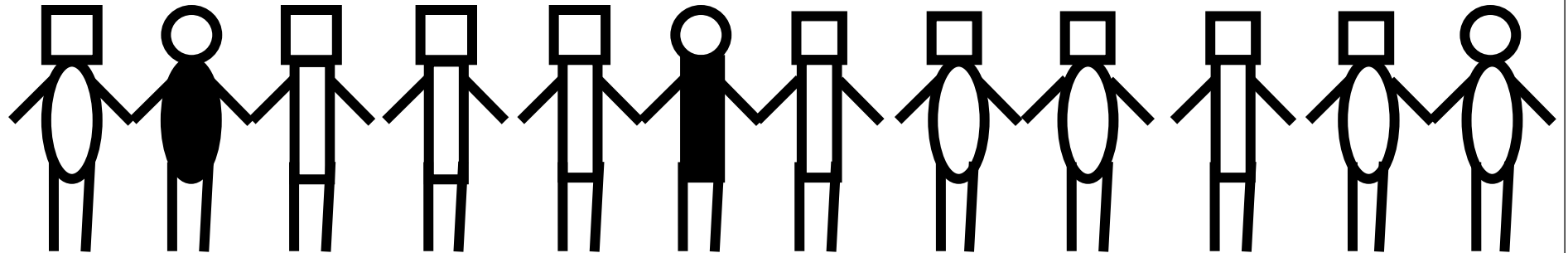
Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	Quad	Red	Quad	Quad	Quad	Red	Quad	Quad	Quad	Quad	Quad	Red
Cor do Corpo	Branco	Preto	Branco	Branco	Branco	Preto	Branco	Branco	Branco	Branco	Branco	Branco
Crédito	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Sim	Não	Sim



Vamos transformar nosso dataset – Fase 2

Cabeça Quadrada = 1 Redonda = 0

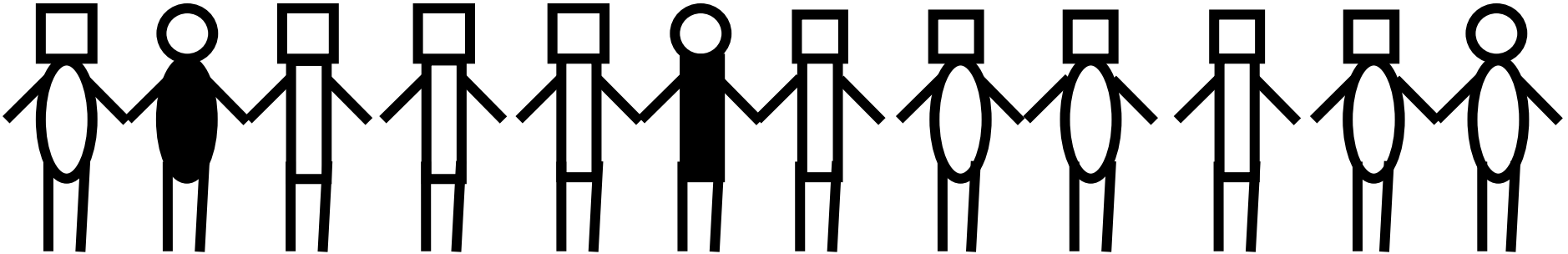
Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	Branco Preto Branco Branco Branco Preto Branco Branco Branco Branco Branco Branco Branco											
Crédito	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Sim	Não	Sim



Vamos transformar nosso dataset – Fase 2

Corpo Branco = 1 Preto = 0

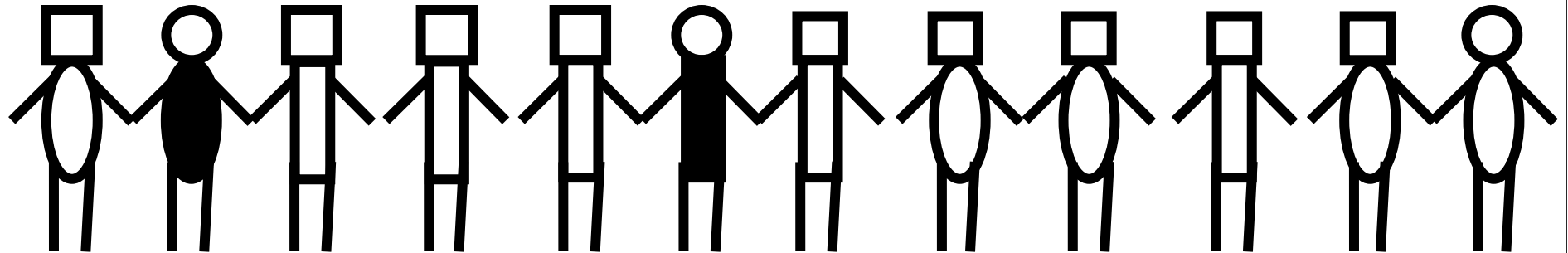
Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	Não	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Sim	Não	Sim



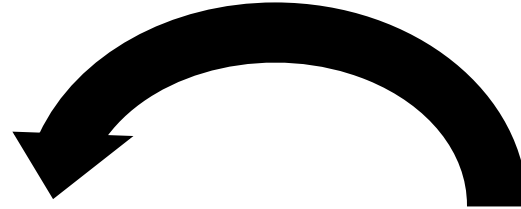
Vamos transformar nosso dataset – Fase 2

Crédito Sim = 1 Não = 0

Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	0	1	1	1	1	0	1	0	0	1	0	1

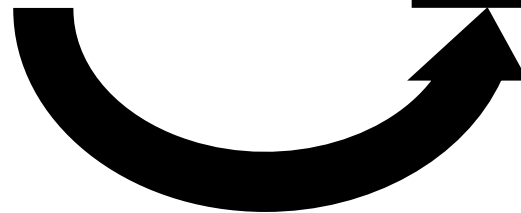


Vamos inverter nossos dados



Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	0	1	1	1	1	0	1	0	0	1	0	1

Instância	Corpo	Cabeça	Cor do Corpo	Crédito
1	1	1	1	0
2	0	0	0	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	0	0	0
7	1	1	1	1
8	0	1	1	0
9	0	1	1	0
10	1	1	1	1
11	0	1	1	0
12	0	0	1	1



Como seriam utilizados os dataset em cada tipo de aprendizado

Aprendizado Supervisionado

Instância	Corpo	Cabeça	Cor do Corpo	Crédito
1	1	1	1	0
2	0	0	0	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	0	0	0
7	1	1	1	1
8	0	1	1	0
9	0	1	1	0
10	1	1	1	1
11	0	1	1	0
12	0	0	1	1

Atributos Preditores

Atributos Alvo

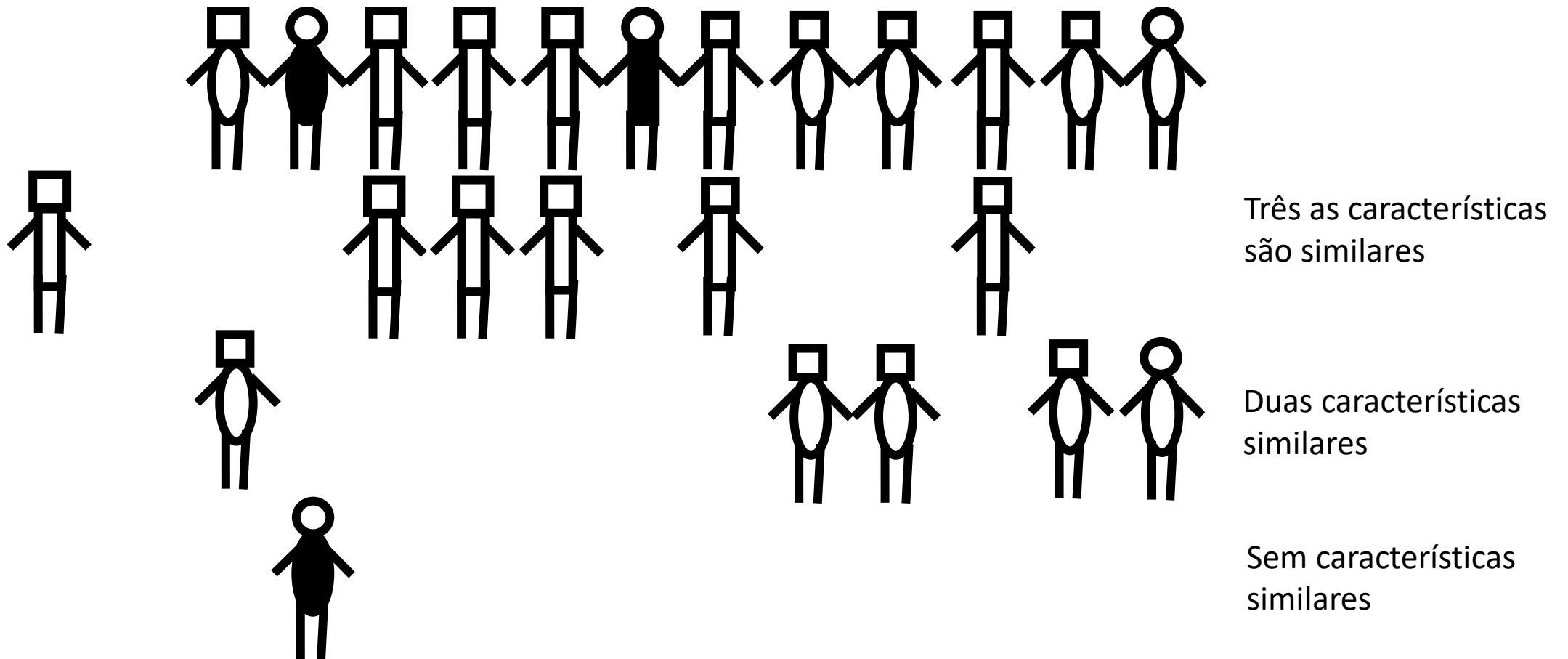
Aprendizado Não Supervisionado

Instância	Corpo	Cabeça	Cor do Corpo	Crédito
1	1	1	1	0
2	0	0	0	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	0	0	0
7	1	1	1	1
8	0	1	1	0
9	0	1	1	0
10	1	1	1	1
11	0	1	1	0
12	0	0	1	1

Atributos Preditores

Um novo conceito SIMILARIDADE

- A idéia de similar está ligada ao conceito de semelhança por exemplo :



Medidas de Similaridade

As medidas de similaridade:

- Fornecem valores numéricos que expressam a “distância” entre dois objetos.

Distância Menores

- Quanto menor o valor desta “distância”, mais semelhantes serão os objetos, e tenderão a ficar no mesmo *cluster*.

Distâncias Maiores

- Quanto maior a “distância”, menos similares serão os objetos e, em consequência, eles deverão estar em grupos distintos.

Medidas de Similaridade

Uma função de distância deve ser tal que:

- não assuma valores negativos (o menor valor é 0);
- ser simétrica (a distância do objeto i ao j tem que ser igual à distância do objeto j ao i);
- forneça o valor 0 quando calculada a distância do objeto a si mesmo ou quando dois objetos são idênticos;
- respeite a desigualdade triangular, (dados 3 objetos, a distância entre dois deles tem que ser menor ou igual a soma das distâncias entre esses dois objetos e o terceiro).

Medidas de Similaridade

Distância Euclidiana

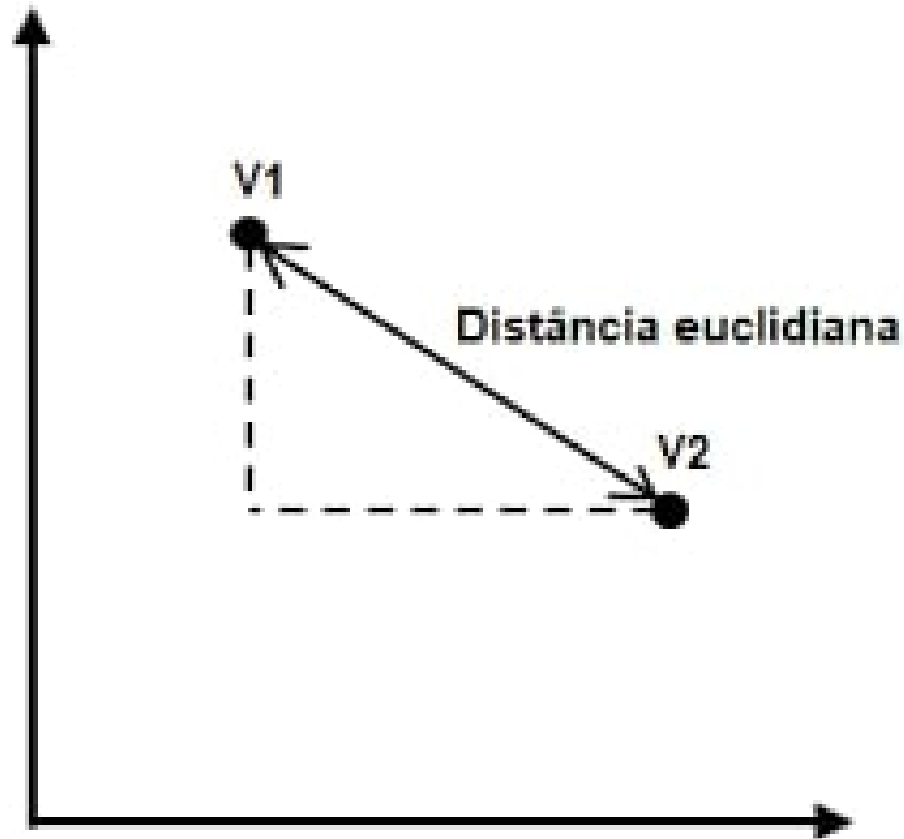
$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

City block (Manhattan, taxicab, L_1 norm, Hamming)

- Um exemplo comum é a distância de Hamming, que é o número de bits que é diferente entre dois vetores binários

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_p - y_p|$$

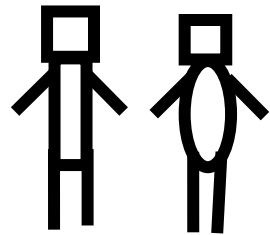
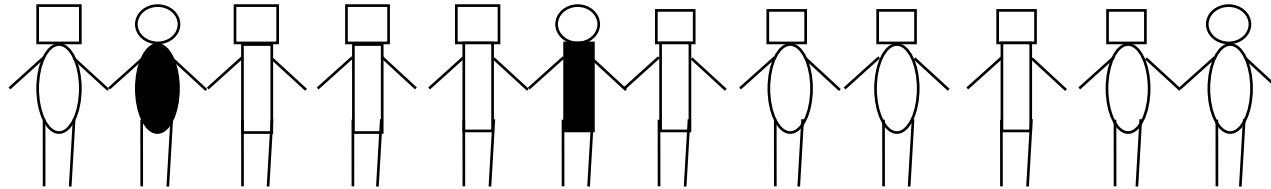
Medidas de Similaridade



- Não há uma **medida** de **similaridade** que sirva para todos os **tipos** de **variáveis** que podem existir numa base de dados.
- **Variáveis numéricas:**
 - A medida que é normalmente usada para computar as dissimilaridades de objetos descritos por variáveis numéricas é a **Distancia Euclidiana**
 - A normalização faz com que todas as variáveis tenham um peso igual. A normalização deve ser efetuada para todos os atributos.

Vamos fazer um exemplo de similaridade calculada por distância

Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	0	1	1	1	1	0	1	0	0	1	0	1

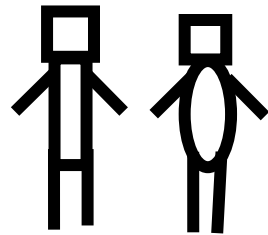
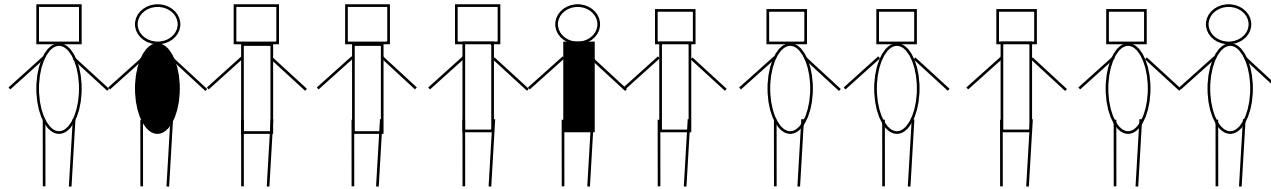


• $d(x, y) =$

$$\sqrt{(corpo_x - corpo_y)^2 + (cabeça_x - cabeça_y)^2 + (credito_x - credito_y)^2 + (cor\ corpo_x - cor\ corpo_y)^2}$$

Vamos fazer um exemplo de similaridade calculada por distância

Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	0	1	1	1	1	0	1	0	0	1	0	1

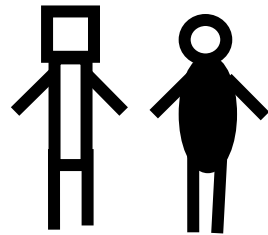
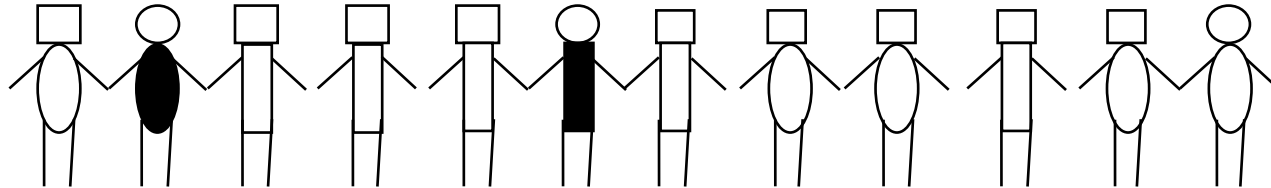


$$\bullet d(x, y) = \sqrt{\begin{aligned} &(1 - 1)^2 + \\ &(1 - 1)^2 + \\ &(1 - 0)^2 + \\ &(1 - 1)^2 \end{aligned}}$$

$$\bullet d(x, y) = 1$$

Vamos fazer um exemplo de similaridade calculada por distância

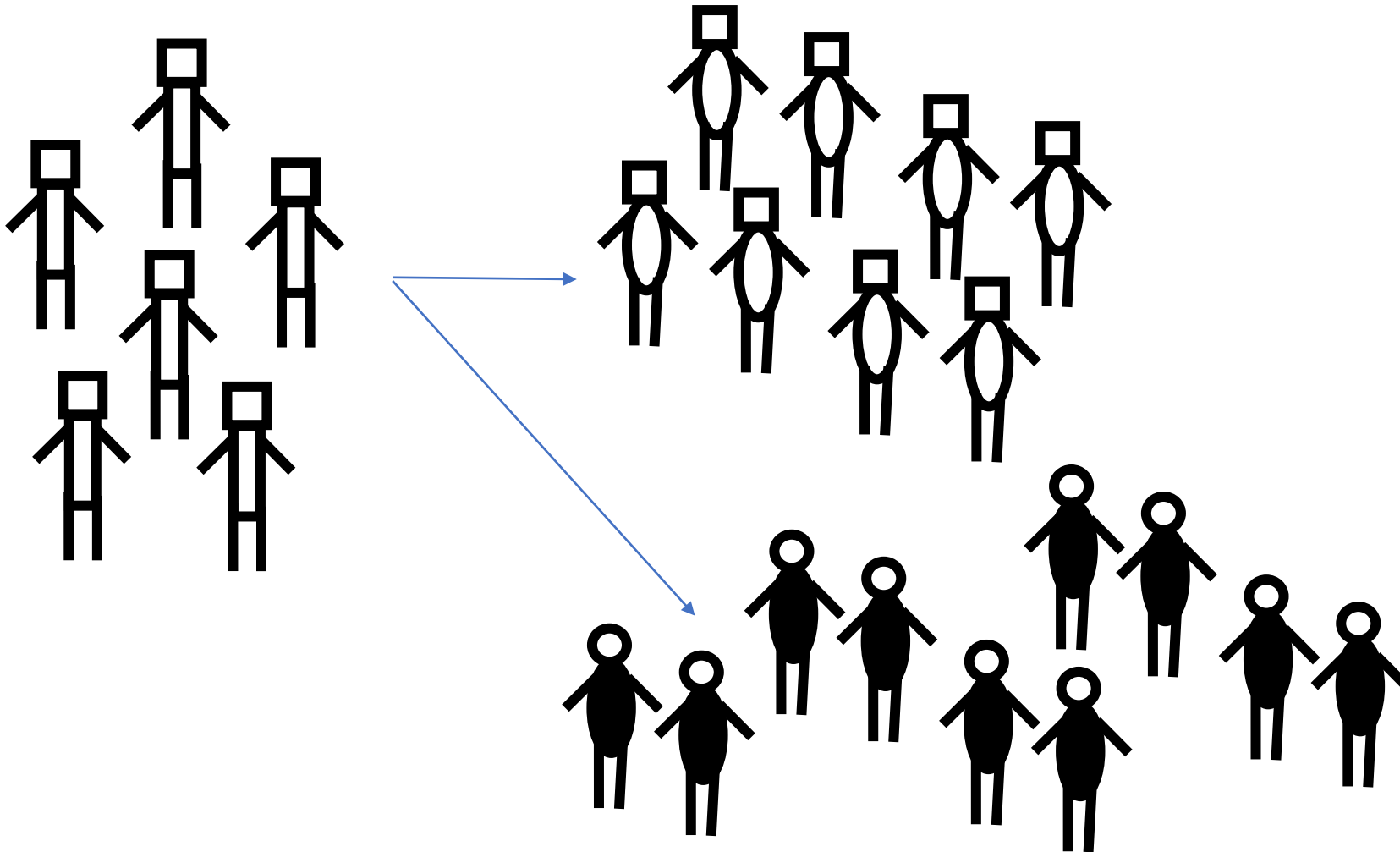
Instância	1	2	3	4	5	6	7	8	9	10	11	12
Corpo	1	0	1	1	1	1	1	0	0	1	0	0
Cabeça	1	0	1	1	1	0	1	1	1	1	1	0
Cor do Corpo	1	0	1	1	1	0	1	1	1	1	1	1
Crédito	0	1	1	1	1	0	1	0	0	1	0	1



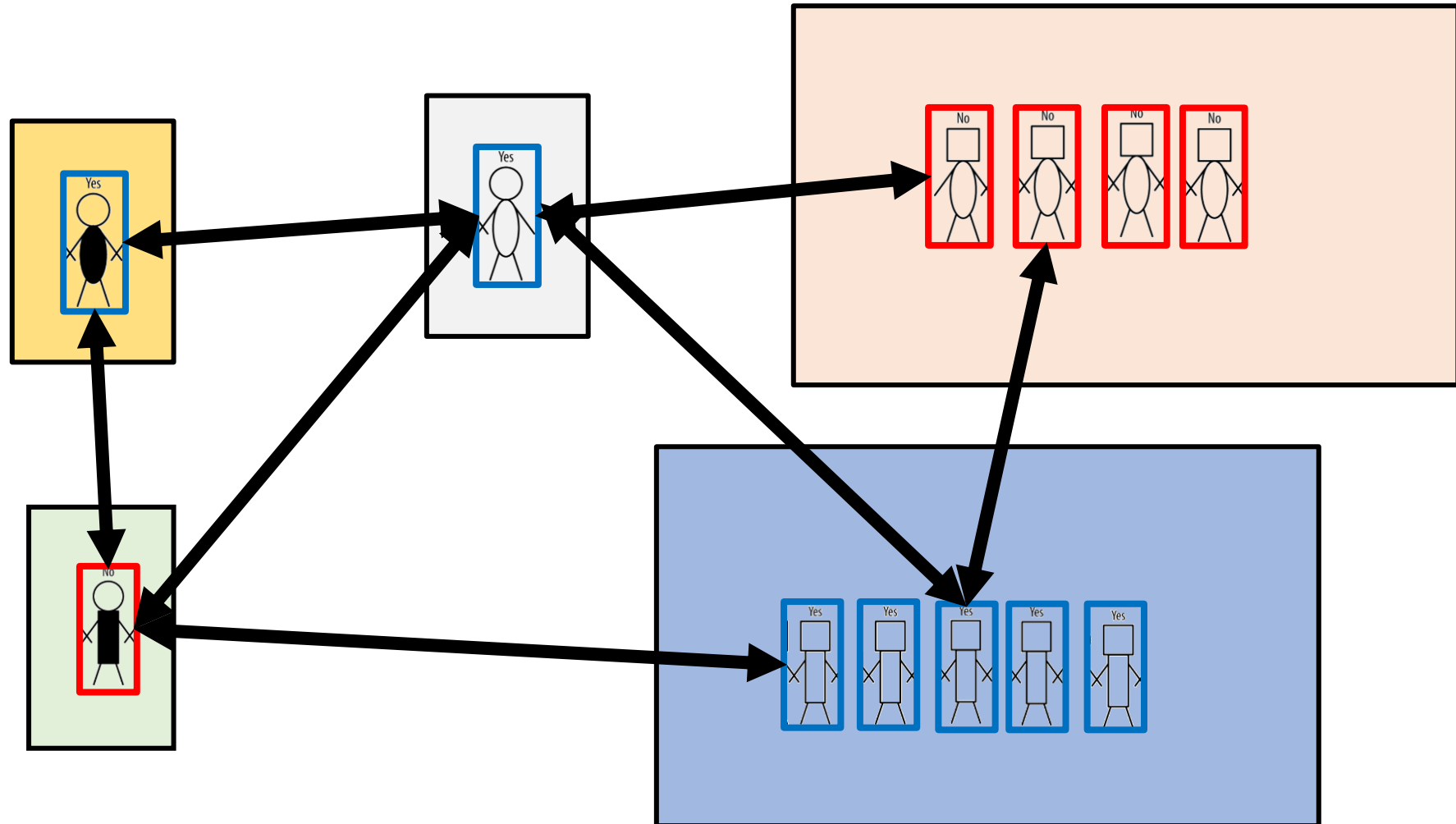
$$\bullet d(x, y) = \sqrt{\begin{matrix} (1 - 0)^2 + \\ (1 - 0)^2 + \\ (1 - 0)^2 + \\ (1 - 1)^2 \end{matrix}}$$

$$\bullet d(x, y) = \sqrt{3}$$

A similaridade permite formar agrupamentos
(clusters)



Distâncias = Similaridade



Kmeans Conceitos

O problema de **Clustering** é descrito como: recebido um conjunto de dados (de objetos), tentar agrupá-los de forma que os elementos que compõem cada grupo sejam mais parecidos entre si do que parecidos com os elementos dos outros grupos.

Em resumo, é colocar os iguais (ou quase iguais) juntos num mesmo grupo e os desiguais em grupos distintos.

O que é formação de agrupamentos (clustering)?

- ❖ Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles
 - ❖ Como agrupar os animais seguintes?
- ❖ Inerentemente é um problema não definido claramente

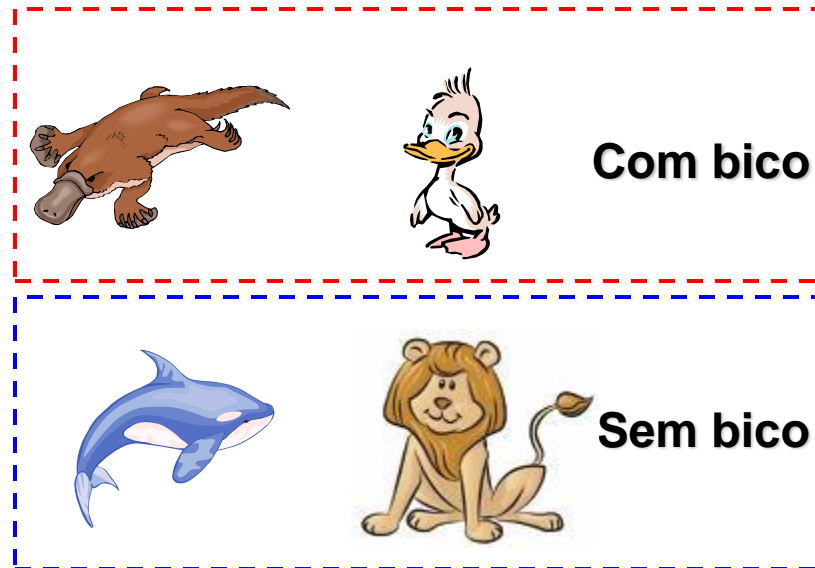


O que é formação de agrupamentos (clustering)?

❖ Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles

❖ Inerentemente é um problema não definido claramente

❖ Como agrupar os animais seguintes?

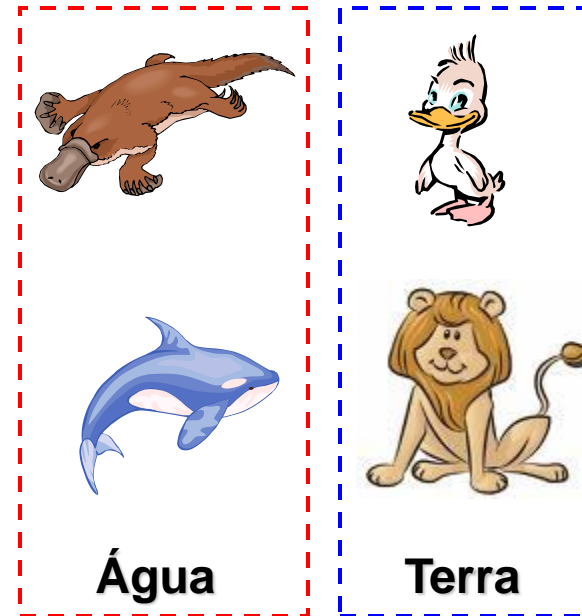


O que é formação de agrupamentos (clustering)?

❖ Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles

❖ Inerentemente é um problema não definido claramente

❖ Como agrupar os animais seguintes?

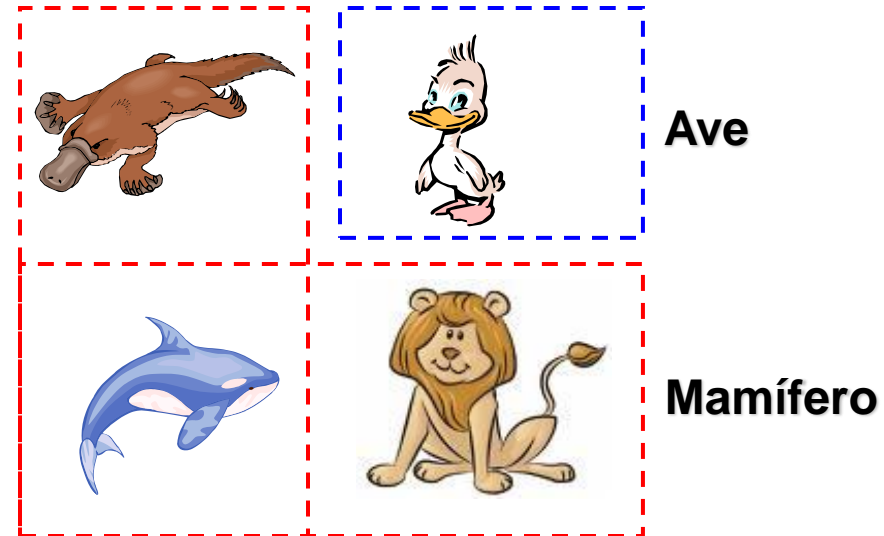


O que é formação de agrupamentos (clustering)?

❖ Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles

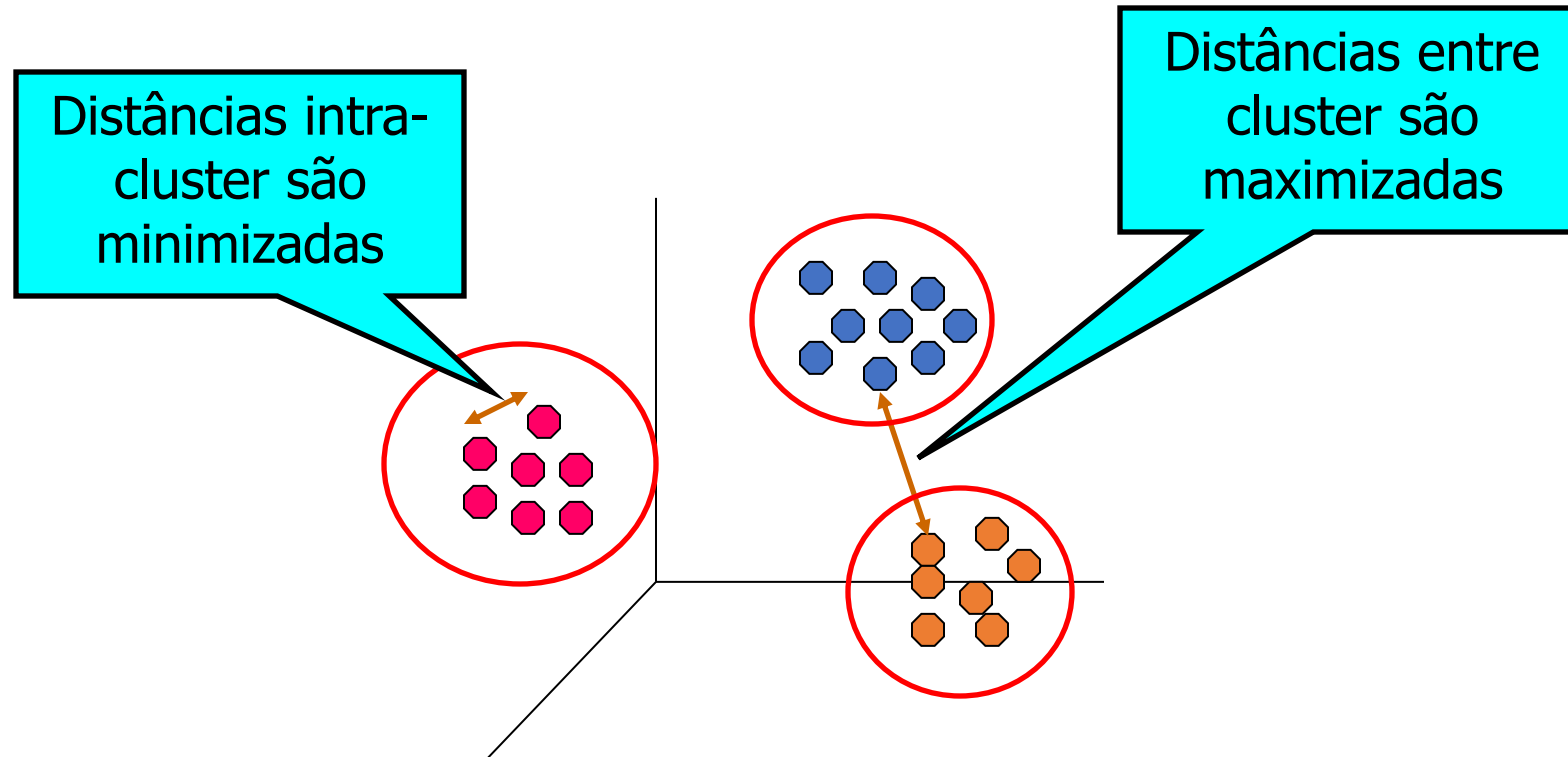
❖ Inerentemente é um problema não definido claramente

❖ Como agrupar os animais seguintes?



O que é formação de agrupamentos (clustering)?

- ❖ Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Aplicações de *clustering*

❖ Entendimento

- Agrupar documentos relacionados, agrupar proteínas com funcionalidades similares, agrupar ações com as mesmas flutuações de preço

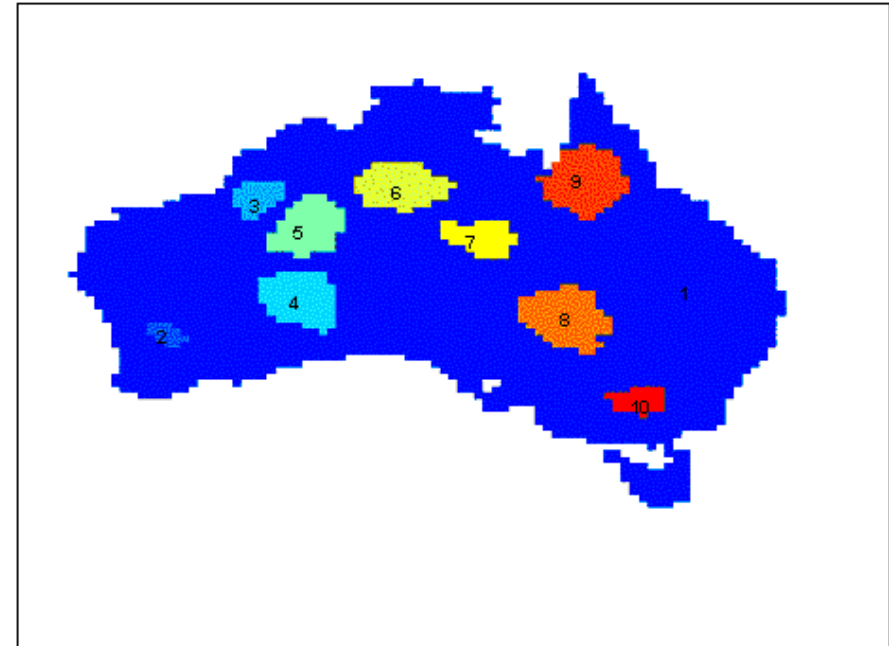
❖ Sumarização

- Reduzir o tamanho de grandes conjuntos de dados

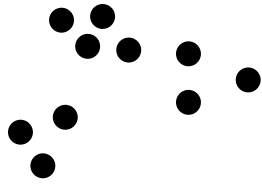
	<i>Discovered Clusters</i>	<i>Industry Group</i>
1	Applied-Matl-DOWN,Bay-Network-DOWN,3-COM-DOWN,Cabletron-Sys-DOWN,CISCO-DOWN,HP-DOWN,DSC-Comm-DOWN,INTEL-DOWN,LSI-Logic-DOWN,Micron-Tech-DOWN,Texas-Inst-DOWN,Tellabs-Inc-DOWN,Natl-Semiconduct-DOWN,Oracl-DOWN,SGL-DOWN,Sun-DOWN	Technology1-DOWN
2	Apple-Comp-DOWN,Autodesk-DOWN,DEC-DOWN,ADV-Micro-Device-DOWN,Andrew-Corp-DOWN,Computer-Assoc-DOWN,Circuit-City-DOWN,Compaq-DOWN,EMC-Corp-DOWN,Gen-Inst-DOWN,Motorola-DOWN,Microsoft-DOWN,Scientific-Atl-DOWN	Technology2-DOWN
3	Fannie-Mae-DOWN,Fed-Home-Loan-DOWN,MBNA-Corp-DOWN,Morgan-Stanley-DOWN	Financial-DOWN
4	Baker-Hughes-UP,Dresser-Inds-UP,Halliburton-HLD-UP,Louisiana-Land-UP,Phillips-Petro-UP,Unocal-UP,Schlumberger-UP	Oil-UP

Agrupando a precipitação na Austrália

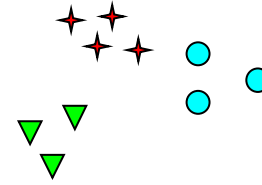
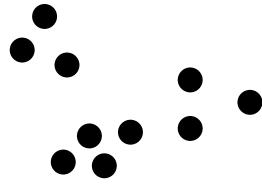
10 Precip Clusters usin SNN Clustering (12 mo. avg, NN = 100)



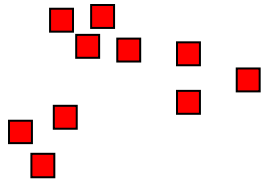
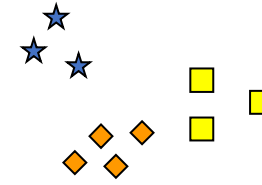
A noção de cluster pode ser ambígua



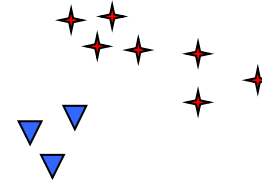
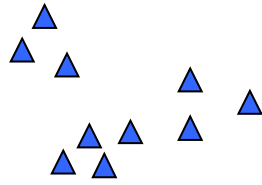
Quantos clusters?



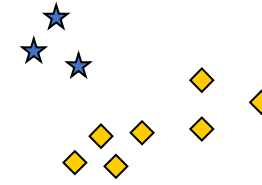
Seis Clusters



Dois Clusters



Quatro Clusters



Dificuldades

▪

- Encontrar o melhor agrupamento para um conjunto de objetos não é uma tarefa simples.
- A não ser que n (número de objetos) e k (número de clusters) sejam extremamente pequenos, visto que o número de partições distintas em que podemos dividir n objetos em k clusters é aproximadamente $k^n/k!$
- Ex. $k=2$ e $n=5$ então são 16 formas de dividir 5 elementos em 2 grupos
- Para agrupar 25 objetos em 5 grupos, existem 2.436.684.974.110.751 maneiras possíveis.
- E se o número de clusters é desconhecido, pois em geral não se sabe qual é o melhor número, precisamos somar todas as partições possíveis para cada número de clusters entre 2 e 5 (desconsiderando um só cluster).

Dificuldades na efetividade de Clustering

Padrões de entrada

- Quase todos os algoritmos de *Clustering* requerem valores para os parâmetros de entrada que são difíceis de determinar, especialmente para conjuntos de dados do mundo real contendo objetos com muitos atributos.

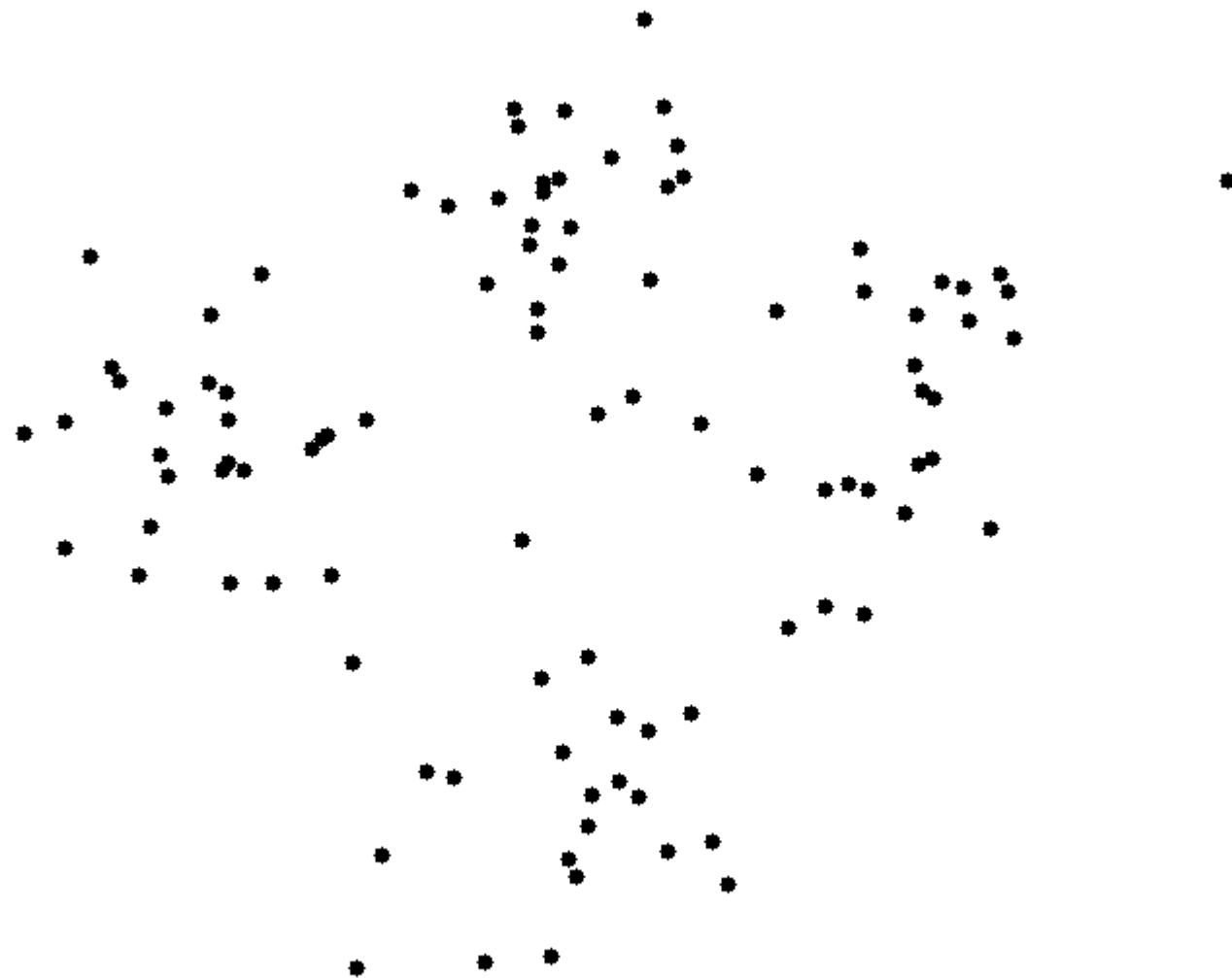
Sensibilidade

- Os algoritmos são muito sensíveis a estes valores de parâmetros, freqüentemente produzindo partições muito diferentes do conjunto de dados mesmo para ajustes de parâmetros significativamente pouco diferentes.

Grande número de atributos (Dimensões)

- Conjuntos de dados reais de alta dimensão (muitos atributos) têm uma distribuição muito ampla o que dificulta a análise.

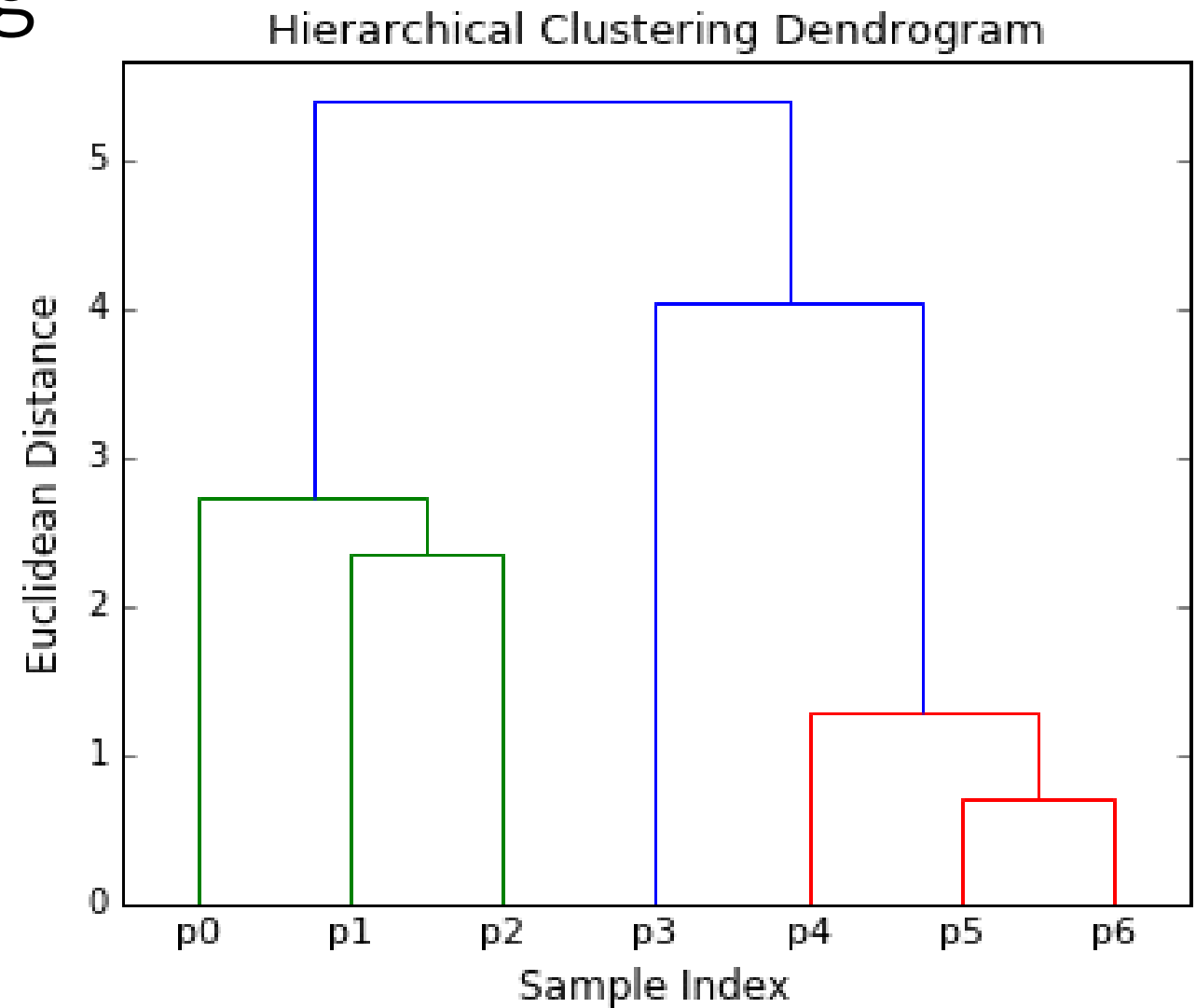
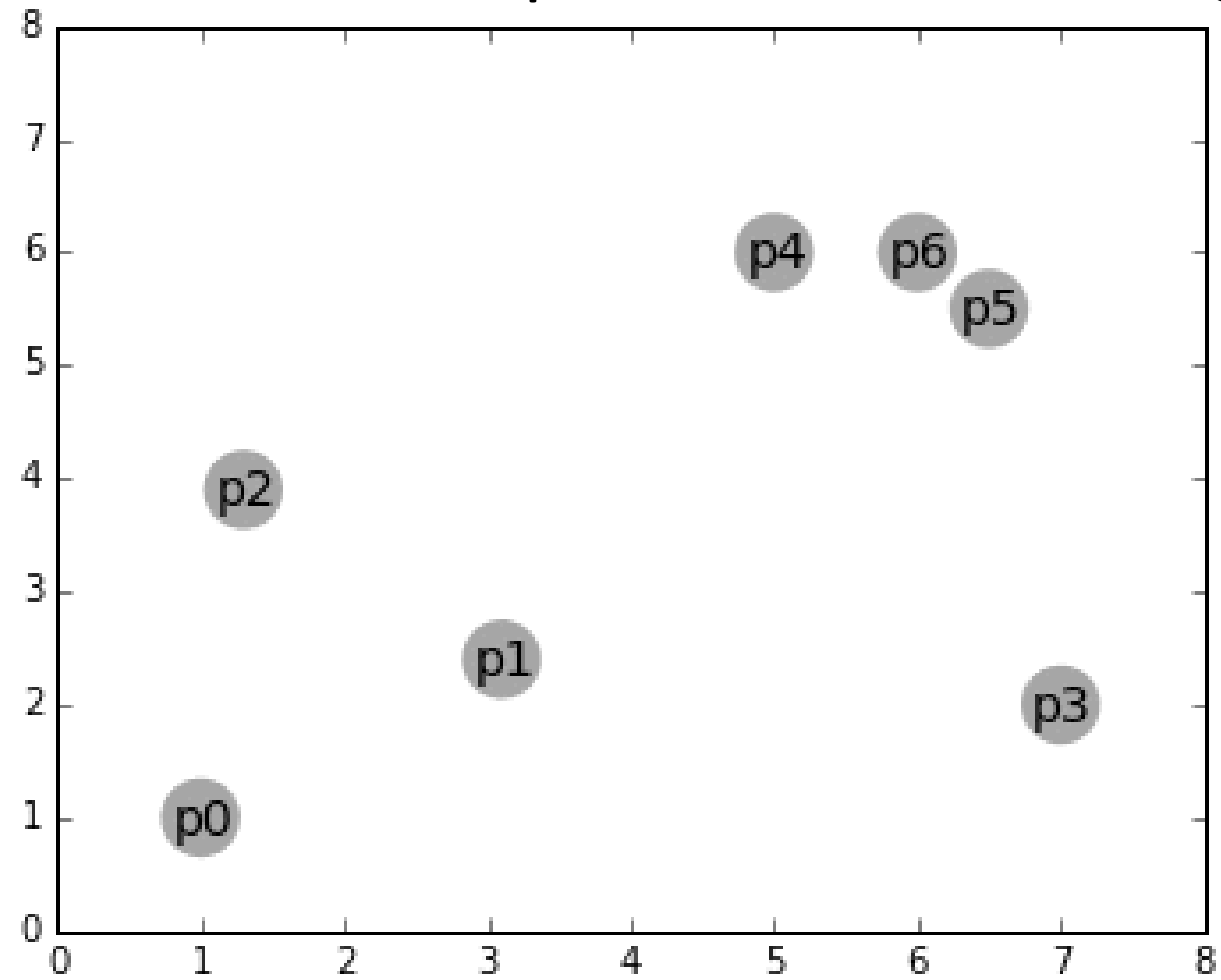
Visualização Medidas de Similaridade



Tipos de algoritmos para Clustering

- Particionamento (K-médias e variantes)
 - Divide os pontos (dados) em conjuntos disjuntos (clusters) tal que cada ponto pertence a um único cluster
- Hierárquica
 - Um conjunto de clusters aninhados organizados como uma árvore
- Baseadas em densidade
 - Encontra clusters baseado na densidade de regiões
- Baseadas em grade (Grid-based)
 - Encontra clusters baseado no número de pontos em cada célula

Hierarchical Clustering

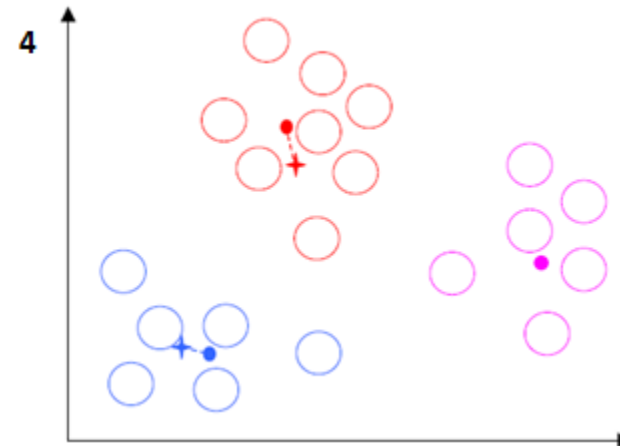
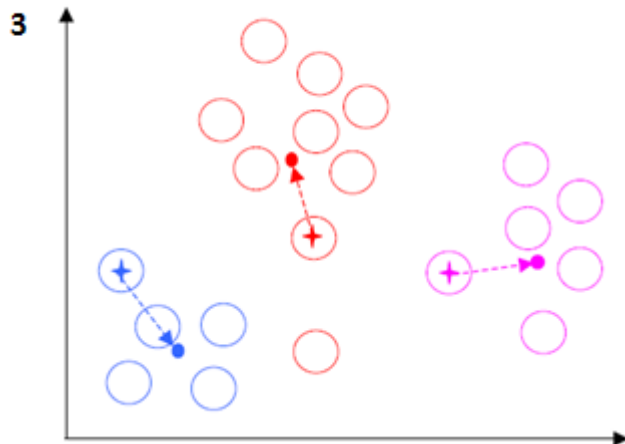
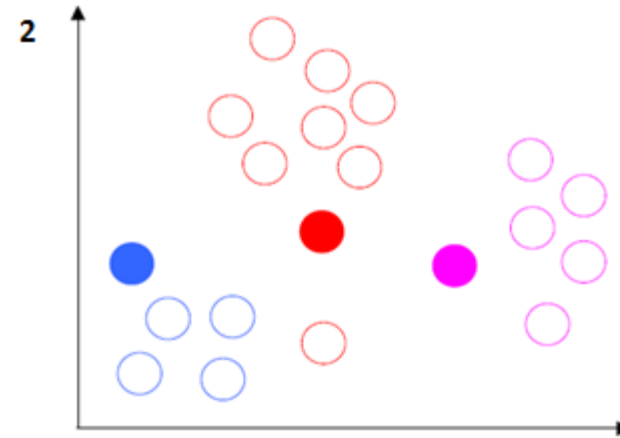
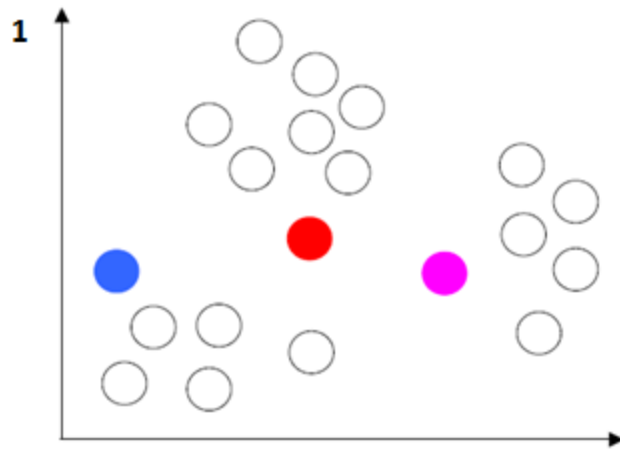


K-médias (K-means)

- Abordagem por particionamento
- Cada cluster está associado a um centróide (ponto central)
- Cada ponto é associado ao cluster cujo centróide está mais próximo
- Número de clusters, K , precisa ser especificado
- O algoritmo básico é bem simples:

-
- 1: Select K points as the initial centroids.
 - 2: **repeat**
 - 3: Form K clusters by assigning all points to the closest centroid.
 - 4: Recompute the centroid of each cluster.
 - 5: **until** The centroids don't change
-

K-médias: exemplo do funcionamento



K-médias: partição

Diagrama de Voronoi – poliedros convexos em torno dos centróides

