

FCT/Unesp – Presidente Prudente
Departamento de Matemática e Computação

Técnicas de Visualização para Dados Multivariados

Parte 3

Prof. Danilo Medeiros Eler
danilo.eler@unesp.br

Sumário

- Parte 1
 - Técnicas Orientadas a Pixel
 - Técnicas Iconográficas
- Parte 2
 - Técnicas Baseadas em Grafos
 - Técnicas Hierárquicas
- Parte 3
 - Técnicas de Projeção Geométrica

Técnicas de Projeção Geométrica

Técnicas de Projeção Geométrica

- Técnicas de Projeção Geométrica procuram transformar o espaço multidimensional dos dados
- Geralmente, utilizam recursos geométricos para representar o relacionamento entre as instâncias ou os atributos
 - Por exemplo, linhas, pontos e plano

Técnicas de Projeção Geométrica

- As técnicas de Projeção Geométrica mais conhecidas são
 - Gráficos de Dispersão
 - Matriz de Gráficos de Dispersão
 - Coordenadas Paralelas
 - RadViz
 - Projeções Multidimensionais
 - Técnicas Baseadas em Força
 - Redução de Dimensionalidade
 - Decomposição Espectral

Gráficos de Dispersão

- Gráficos de Dispersão (scatterplots) são uma das técnicas de visualização mais antigas e mais utilizadas para análise de dados
- Podem representar os dados por meio de
 - Um subconjunto de dimensões: o usuário pode selecionar alguns atributos ou desenvolver métodos para encontrar dimensões que contenham informações úteis

Gráficos de Dispersão

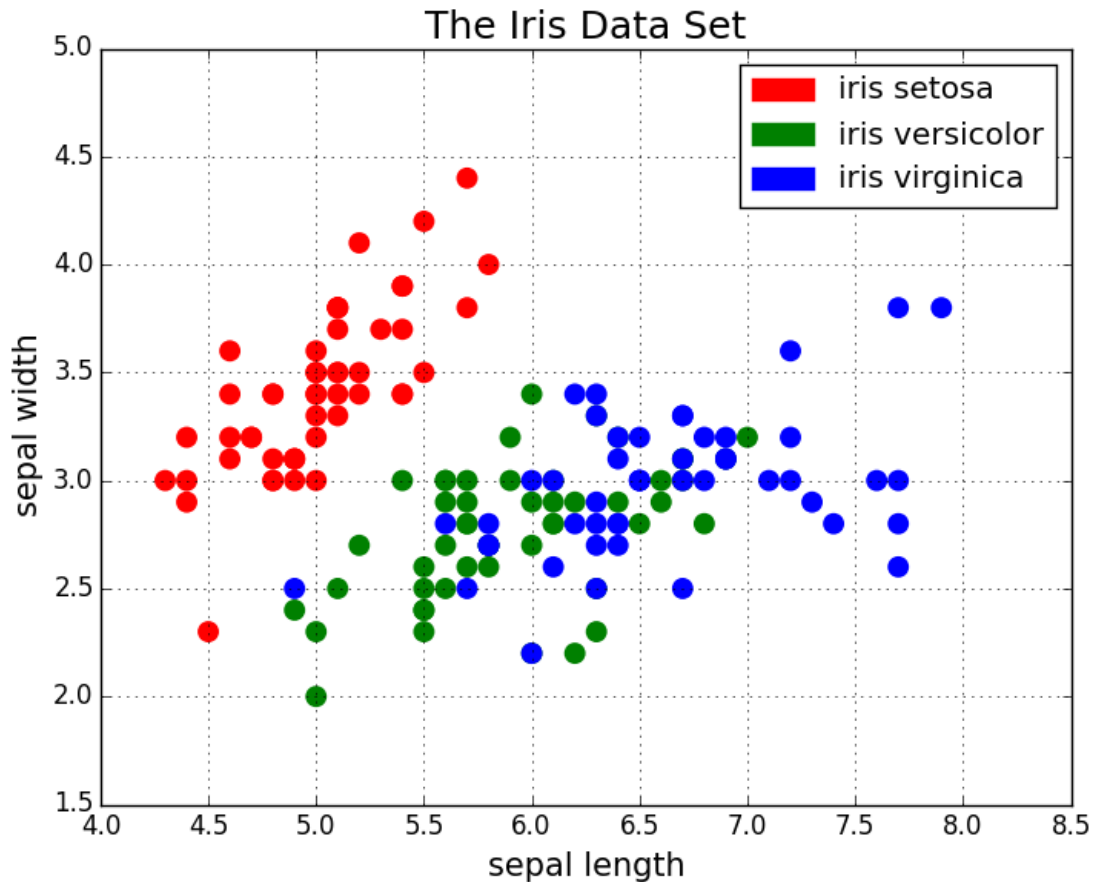
- Gráficos de Dispersão (scatterplots) são uma das técnicas de visualização mais antigas e mais utilizadas para análise de dados
- Podem representar os dados por meio de
 - Redução de dimensionalidade: técnicas de projeção podem se utilizadas para transformar o espaço multidimensional

Gráficos de Dispersão

- Gráficos de Dispersão (scatterplots) são uma das técnicas de visualização mais antigas e mais utilizadas para análise de dados
- Podem representar os dados por meio de
 - Dimensão embutida: outros atributos podem ser apresentados por meio de cor, tamanho e forma
 - Múltiplos gráficos: essa abordagem permite a combinação de diferentes dimensões
 - Uma estratégia conhecida é a Matriz de Gráficos de Dispersão

Gráficos de Dispersão

■ Exemplo para o conjunto Iris

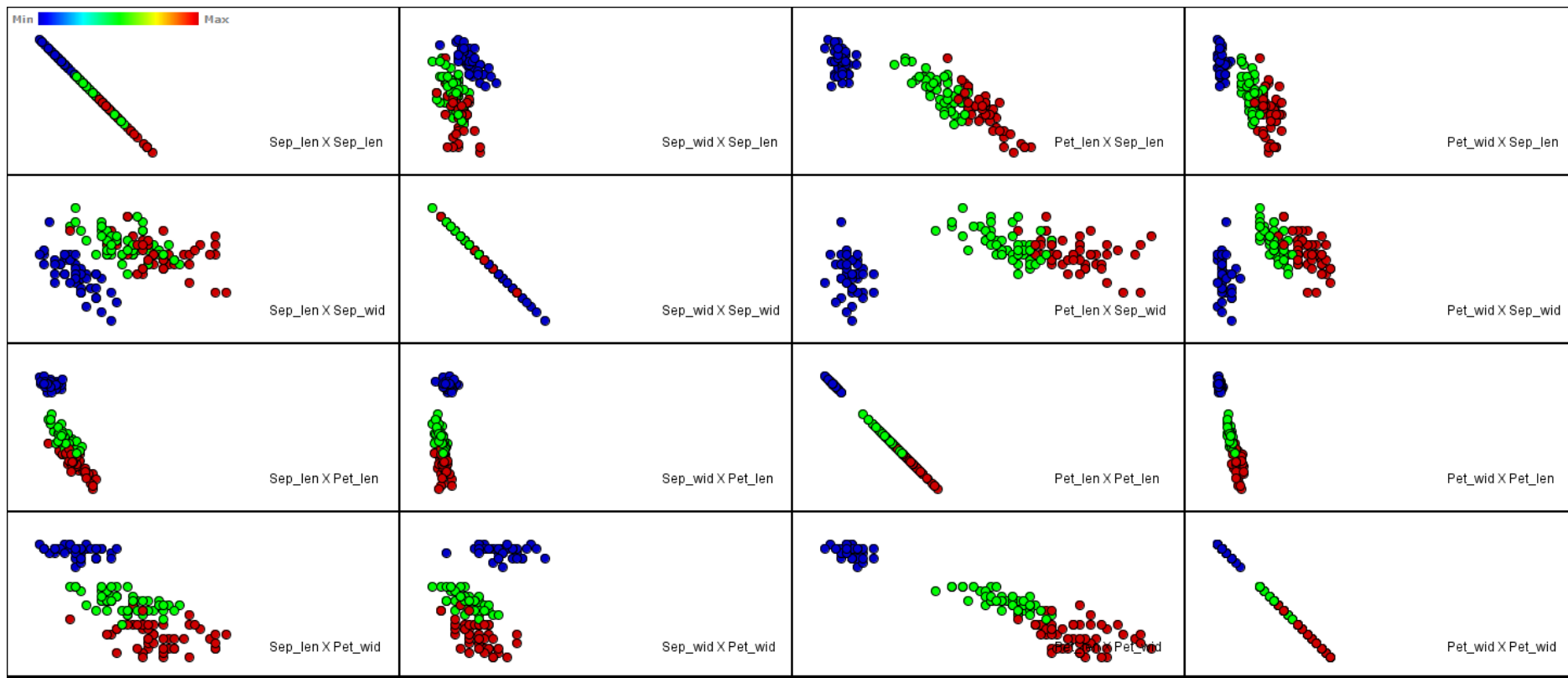


Matriz de Gráficos de Dispersão

- Matriz de Gráficos de Dispersão apresenta diversos Gráficos de Dispersão, permitindo determinar a correlação entre dois atributos
- Além disso, é possível observar o relacionamento entre as instâncias por meio da combinação dos diversos atributos dos dados

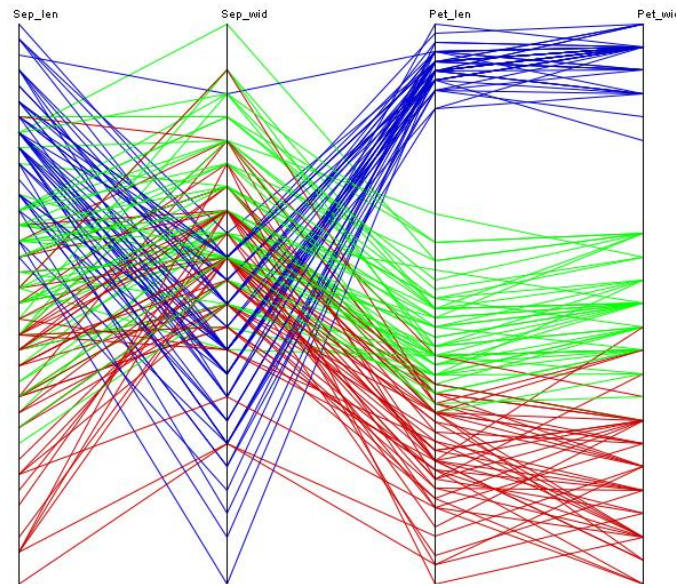
Matriz de Gráficos de Dispersão

■ Exemplo para o conjunto Iris



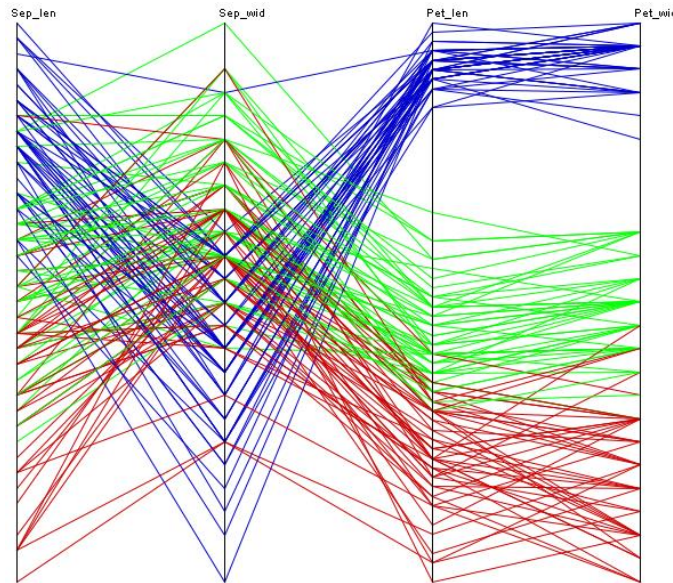
Coordenadas Paralelas

- Coordenadas Paralelas (Inselberg, 1985) é uma das técnicas de visualização mais conhecidas e estudadas
- Ela ainda é foco de estudo de muitos pesquisadores e de diversas variações para melhorar a análise de dados multivariada



Coordenadas Paralelas

- A ideia básica da técnica é que eixos paralelos são posicionados para representar as dimensões do conjunto de dados
- Em seguida, cada instância é desenhada como uma linha cruzando cada eixo na posição proporcional ao valor de sua dimensão



Coordenadas Paralelas

- Exemplo para o conjunto de dados Iris



Iris setosa

sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8



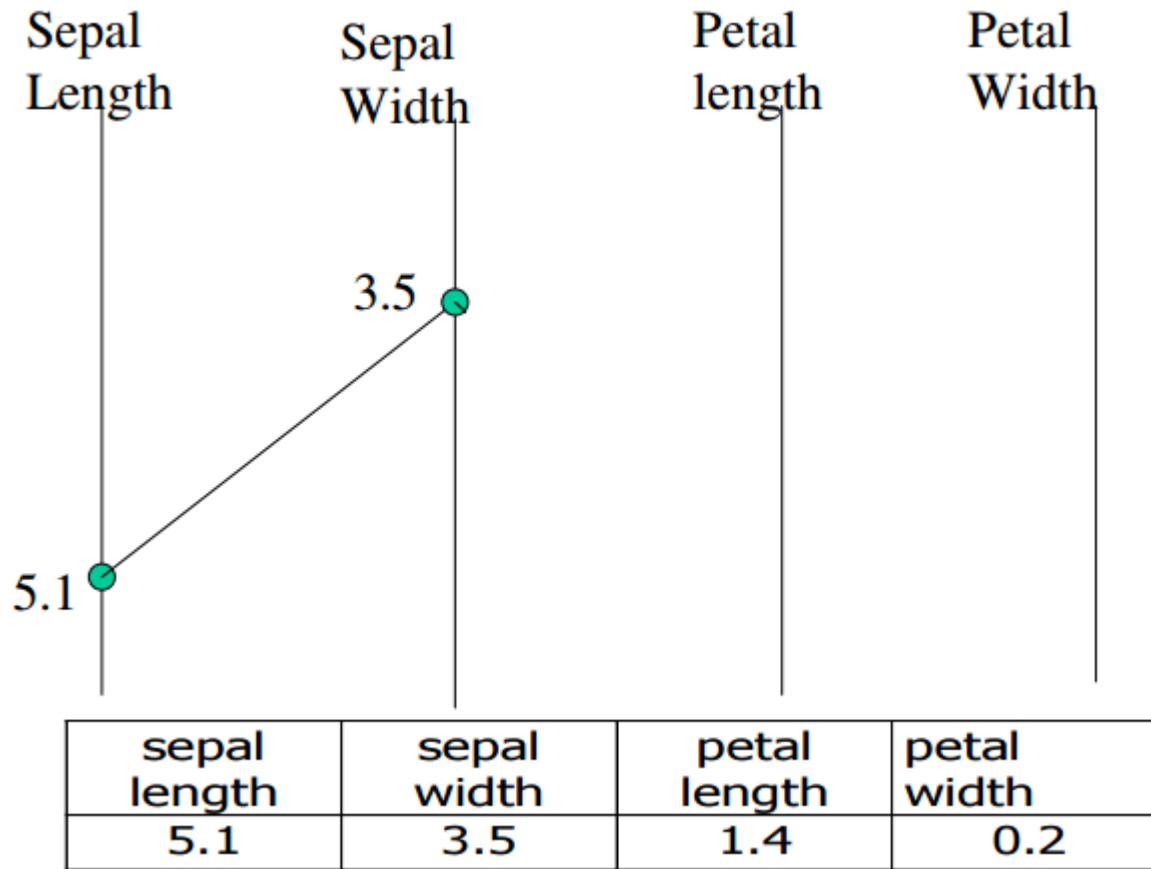
Iris versicolor



Iris virginica

Coordenadas Paralelas

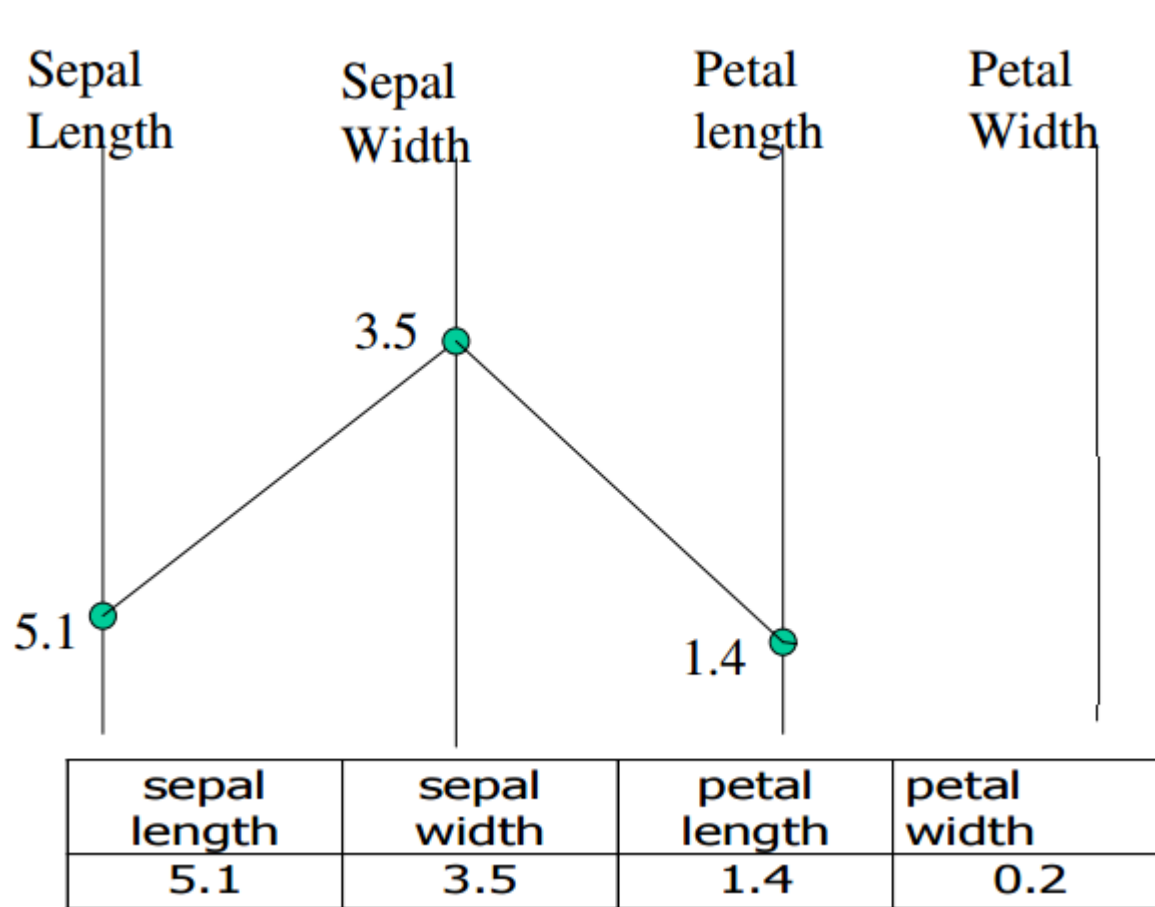
■ Exemplo do conjunto de dados Iris



sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8

Coordenadas Paralelas

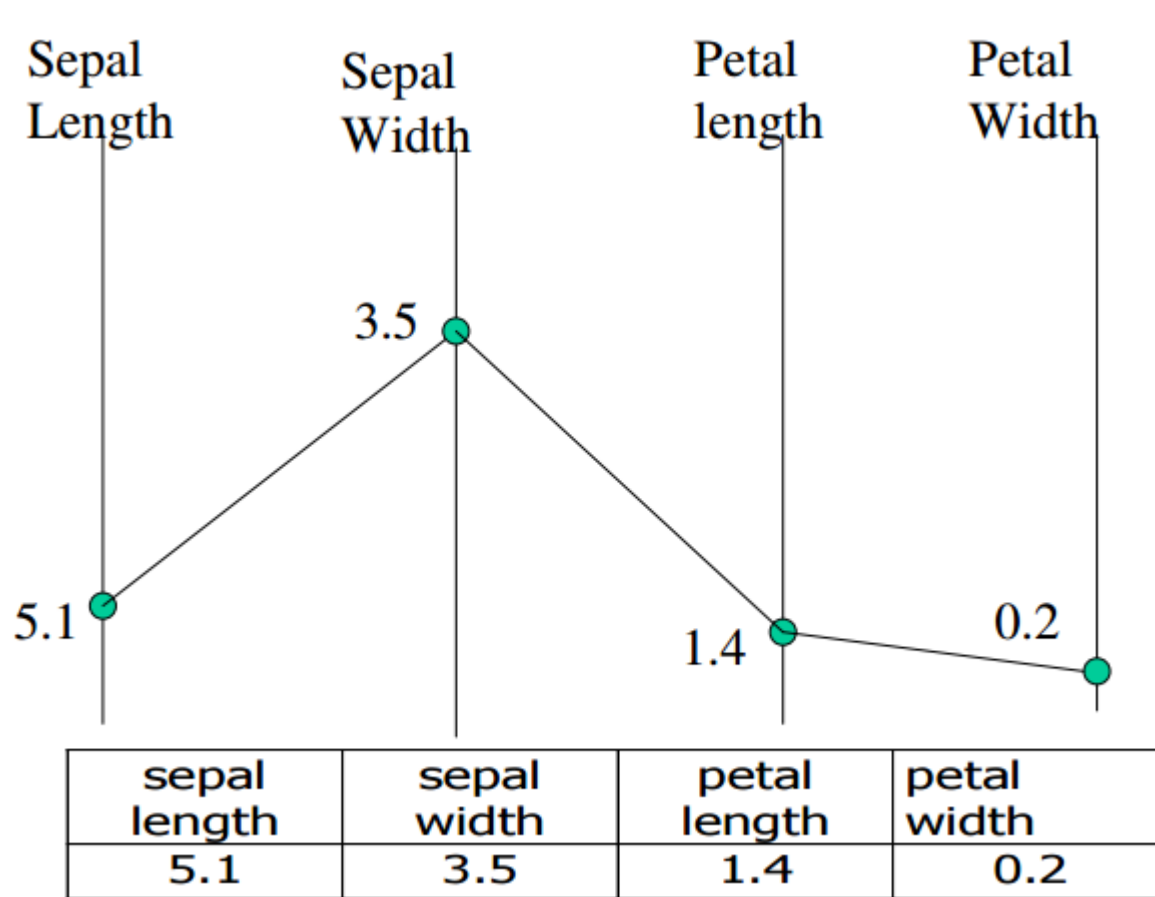
■ Exemplo do conjunto de dados Iris



sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8

Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris

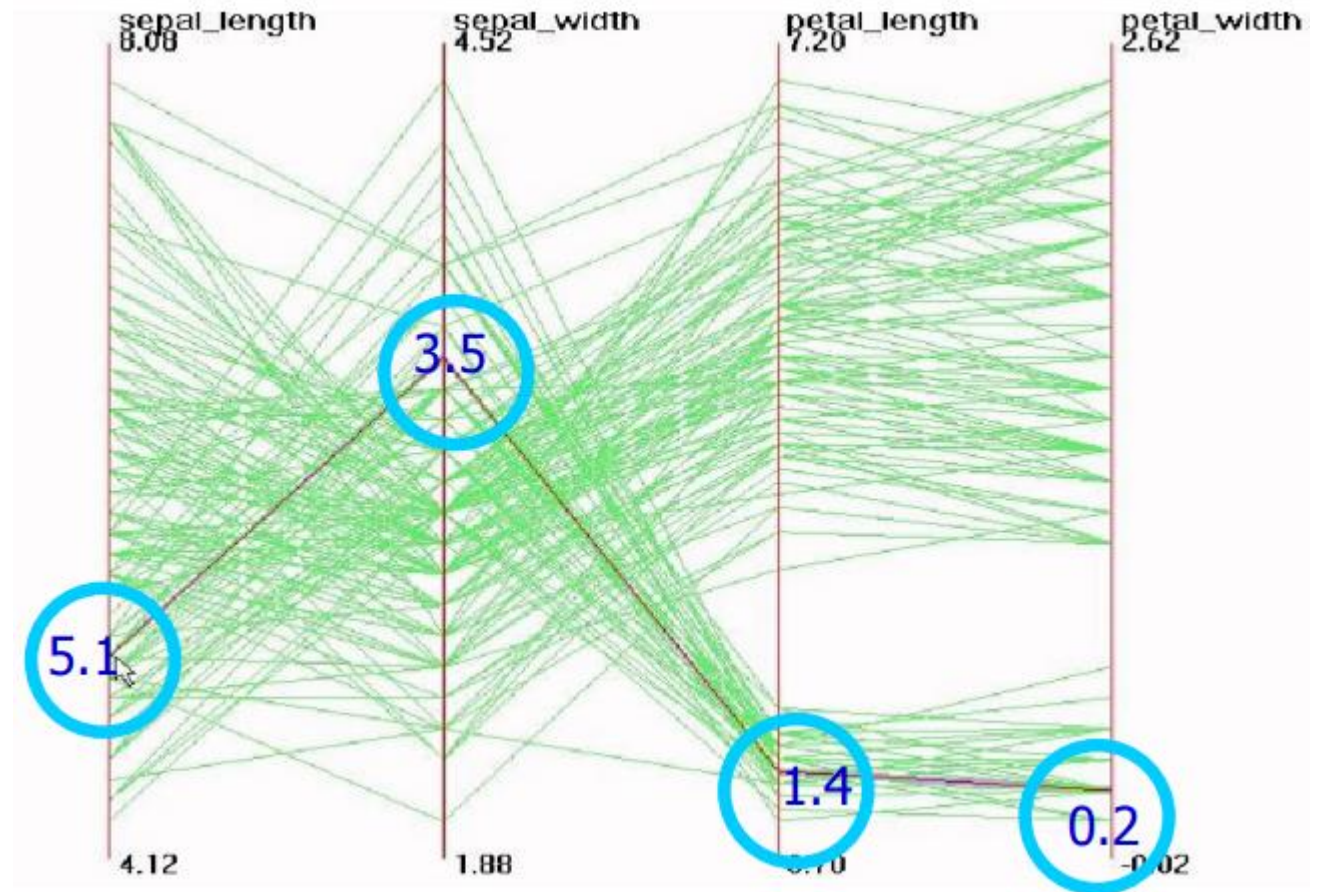


sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8

Coordenadas Paralelas

- Exemplo do conjunto de dados Iris

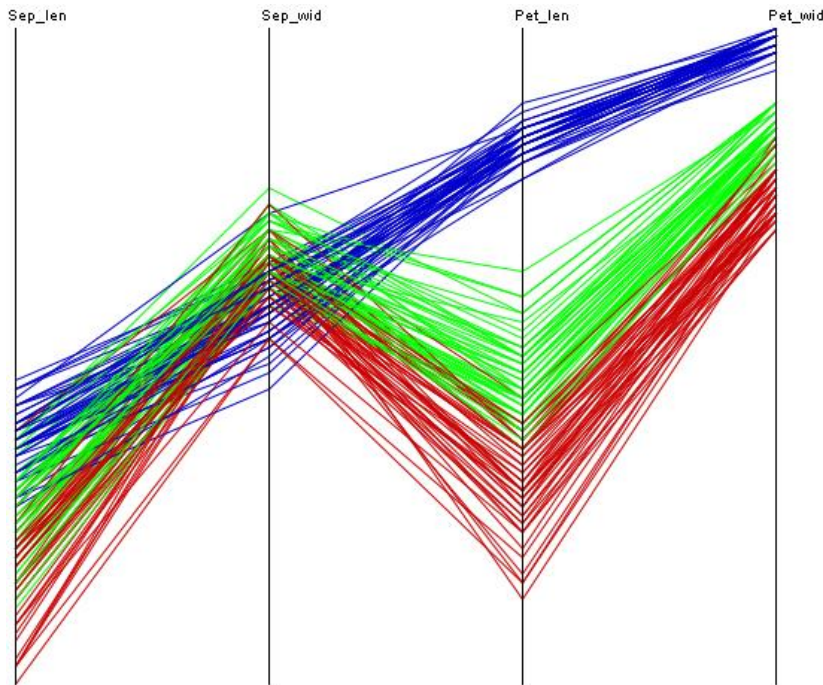
sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8



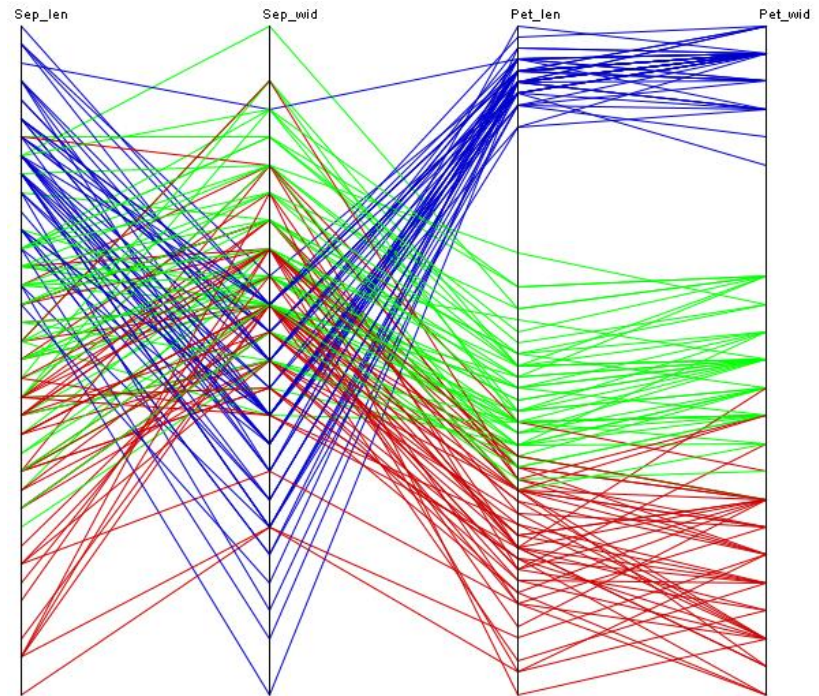
Coordenadas Paralelas

- Exemplo do conjunto de dados Iris

Normalização por todos os atributos

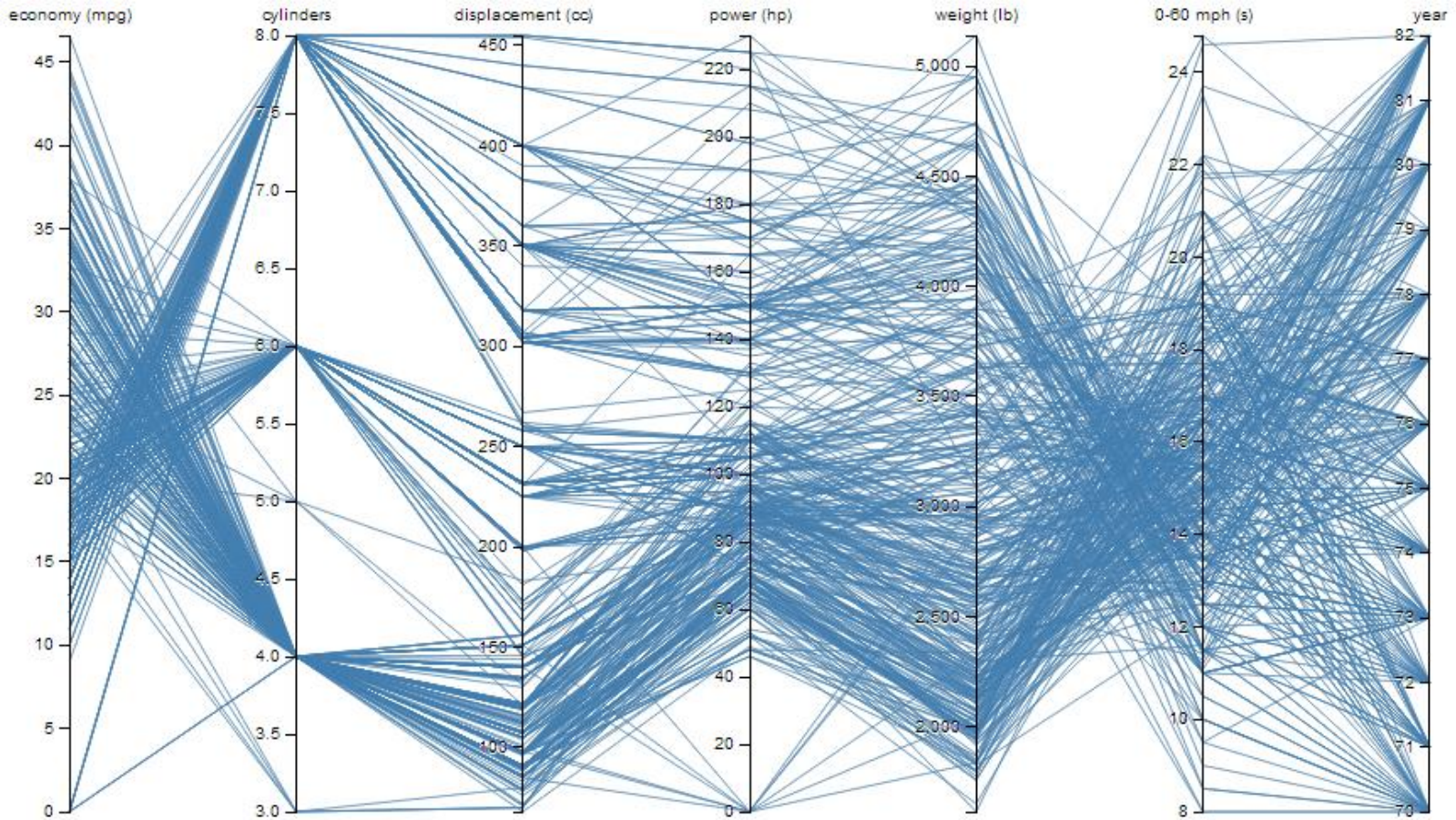


Normalização por atributo



Coordenadas Paralelas

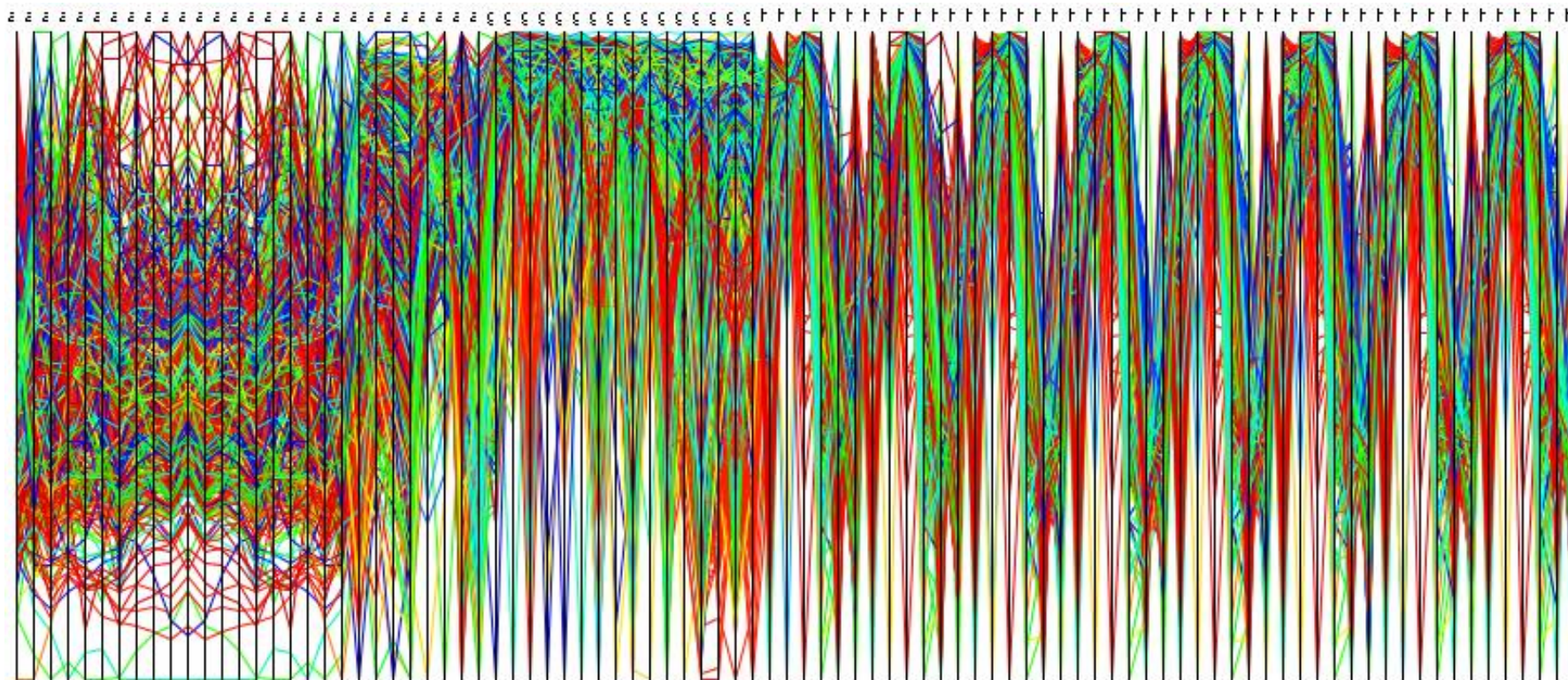
■ Conjunto de dados de carros



<http://i.stack.imgur.com/bLzAX.png>

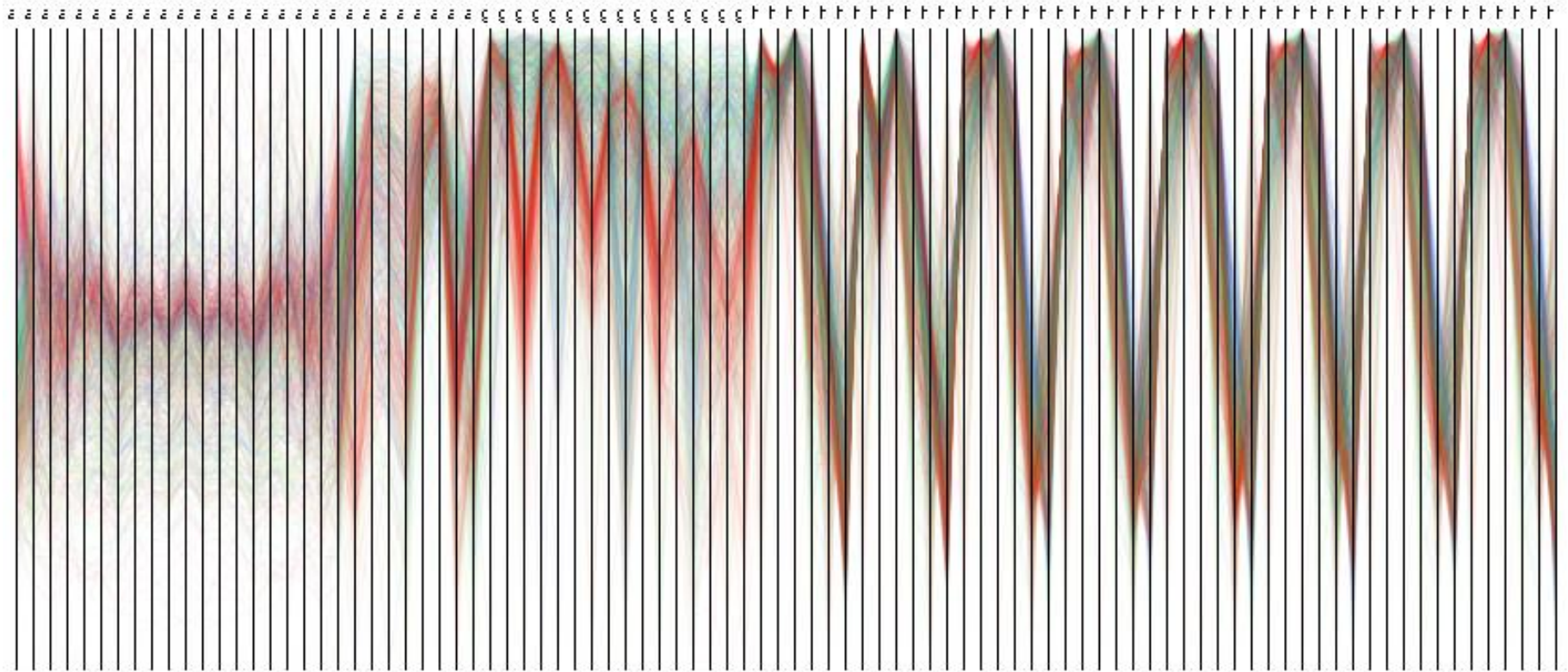
Coordenadas Paralelas

- Conjunto ImageClef 2006
 - 1000 instâncias e 96 atributos



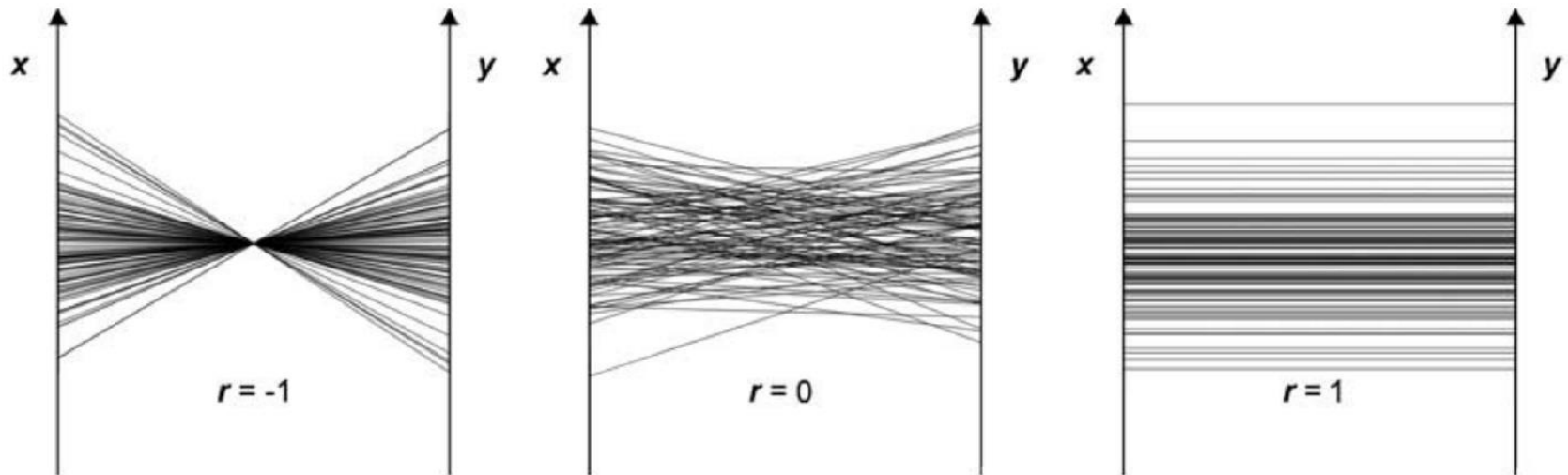
Coordenadas Paralelas

- Conjunto ImageClef 2006
 - 1000 instâncias e 96 atributos
 - Utilização de transparência



Coordenadas Paralelas

- Alguns sistemas permitem que os eixos sejam automaticamente ordenados
 - Por exemplo, baseado na correlação das características

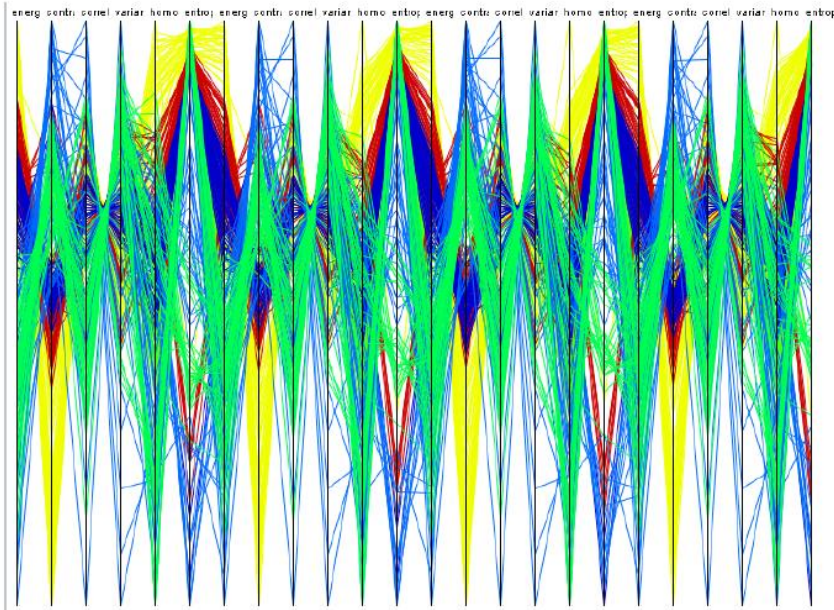


r = coeficiente de correlação

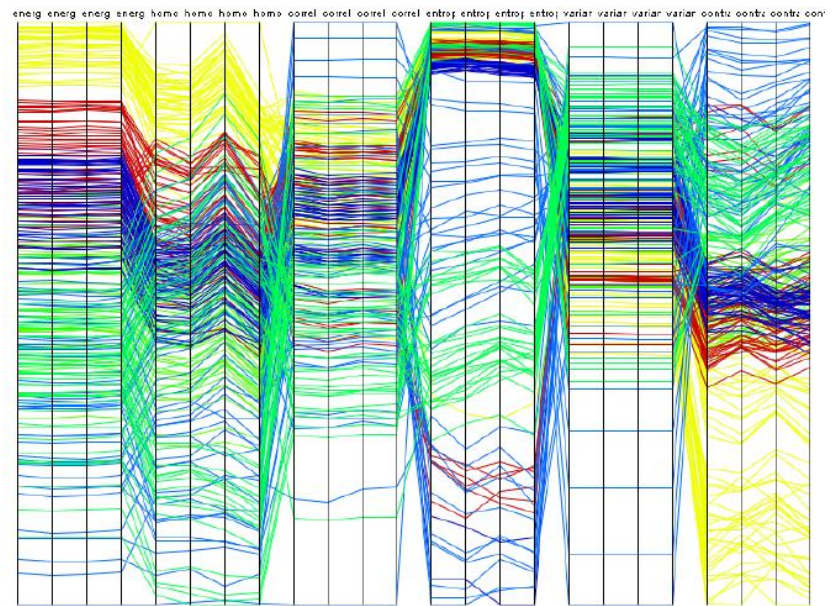
Coordenadas Paralelas

- Conjunto de 270 imagens de tomografia com características computadas por Matrizes de Covariância

Sem ordenação dos eixos

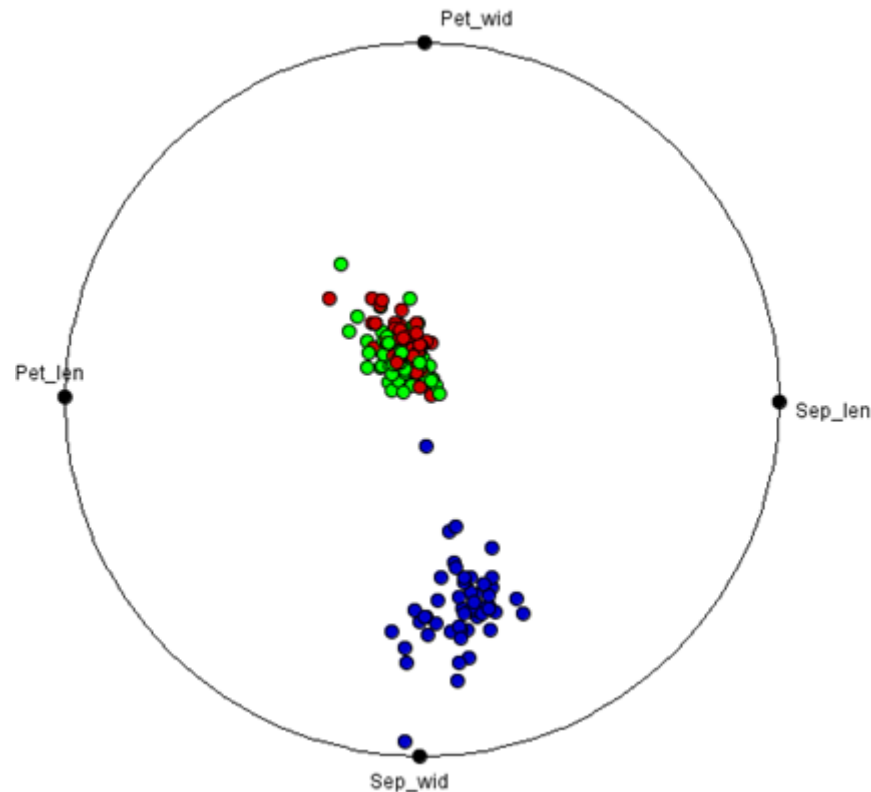


Ordenação baseada na correlação



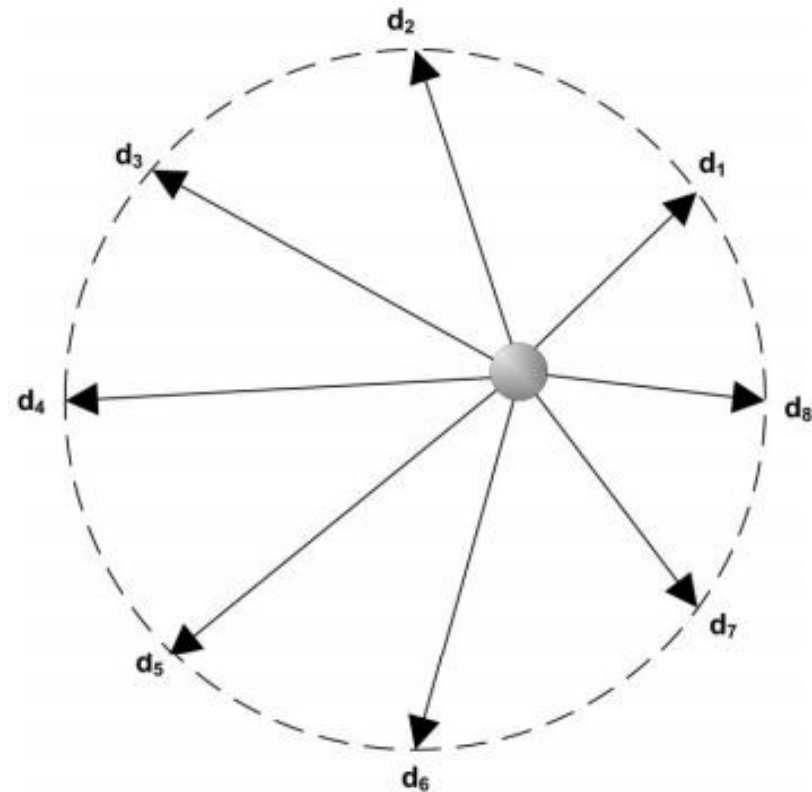
RadViz

- RadViz é uma técnica dirigida por forças baseada na lei de equilíbrio de Hooke



RadViz

- Para um conjunto de dados N-dimensional, N âncoras são posicionadas em uma circunferência
- Elas representam o final de molas conectadas para cada instância



RadViz

- Para simplificar os cálculos e melhorar o entendimento do algoritmo, essas âncoras são posicionadas em um círculo de raio 1.0 centrado na origem
 - Assim, dado uma instância $D = (d_0, d_1, \dots, d_{N-1})$ e um conjunto de vetores unitários A , onde A_j representa a j -ésima âncora, o ponto p que representa a posição da instância na circunferência é dado por

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

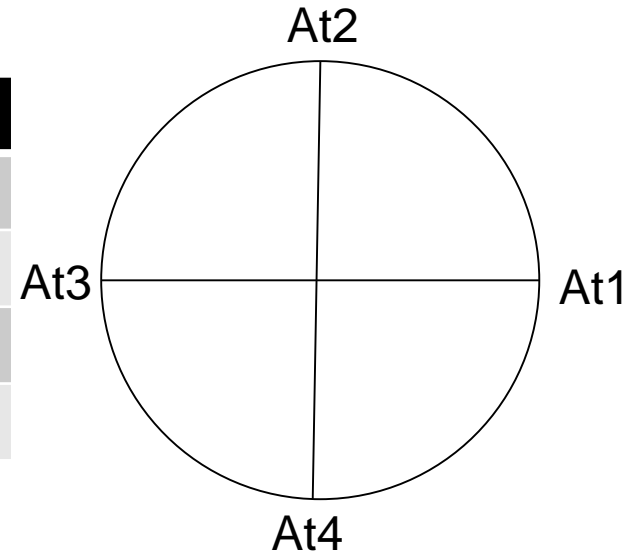
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$Ax = [1, 0, -1, 0]$$

$$Ay = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



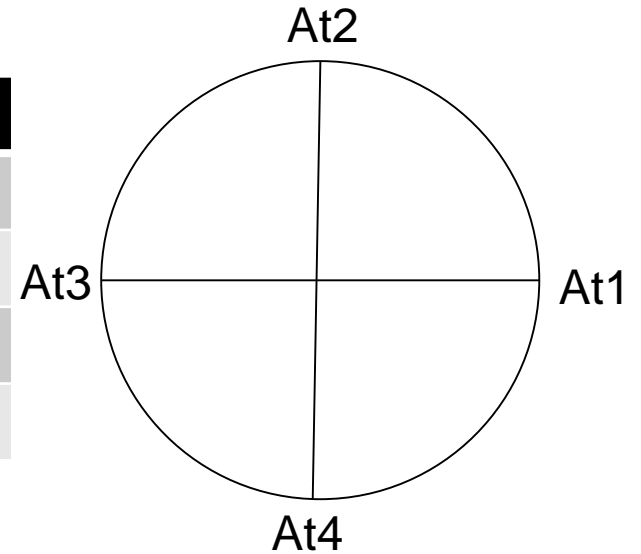
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$Ax = [1, 0, -1, 0]$$

$$Ay = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 10 * 0 + 10 * (-1) + 10 * 0}{10 + 10 + 10 + 10} = \frac{10 - 10}{40} = 0$$

$$P_y = \frac{10 * 0 + 10 * 1 + 10 * 0 + 10 * (-1)}{10 + 10 + 10 + 10} = \frac{10 - 10}{40} = 0$$

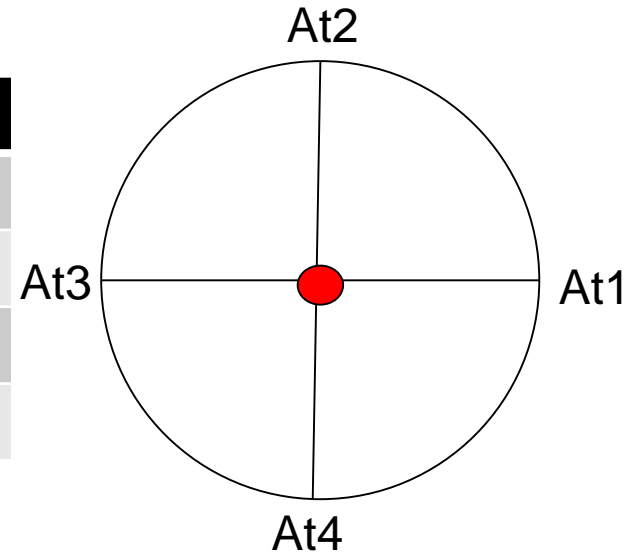
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$Ax = [1, 0, -1, 0]$$

$$Ay = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 10 * 0 + 10 * (-1) + 10 * 0}{10 + 10 + 10 + 10} = \frac{10 - 10}{40} = 0$$

$$P_y = \frac{10 * 0 + 10 * 1 + 10 * 0 + 10 * (-1)}{10 + 10 + 10 + 10} = \frac{10 - 10}{40} = 0$$

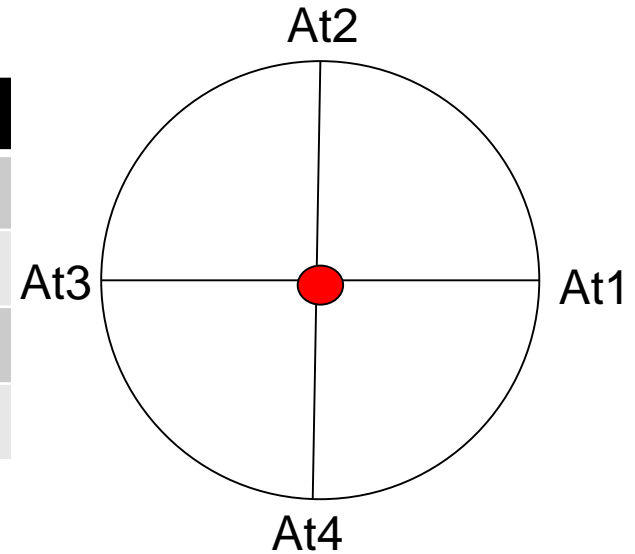
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$A_x = [1, 0, -1, 0]$$

$$A_y = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 0 * 0 + (10) * (-1) + 0 * 0}{10 + 0 + (10) + 0} = \frac{10 - 10}{20} = \frac{0}{20} = 0$$

$$P_y = \frac{10 * 0 + 0 * 1 + (10) * 0 + 0 * (-1)}{10 + 0 + (10) + 0} = \frac{0 + 0}{20} = 0$$

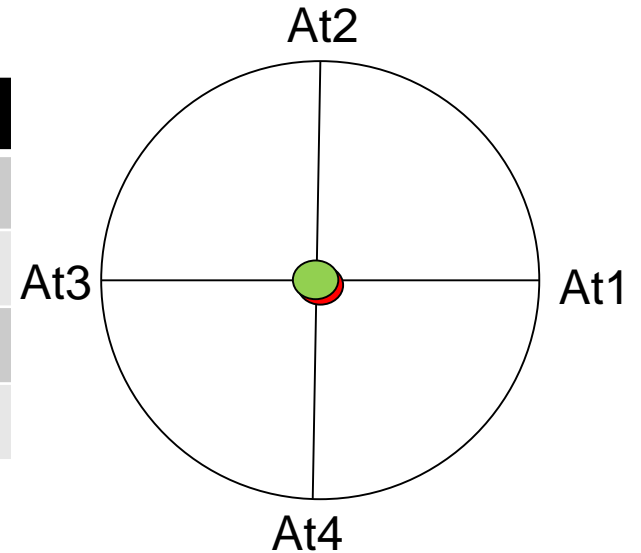
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$A_x = [1, 0, -1, 0]$$

$$A_y = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 0 * 0 + (10) * (-1) + 0 * 0}{10 + 0 + (10) + 0} = \frac{10 - 10}{20} = \frac{0}{20} = 0$$

$$P_y = \frac{10 * 0 + 0 * 1 + (10) * 0 + 0 * (-1)}{10 + 0 + (10) + 0} = \frac{0 + 0}{20} = 0$$

Observação: o círculo verde foi deslocado para vermos a sobreposição, mas na prática não veríamos o círculo vermelho

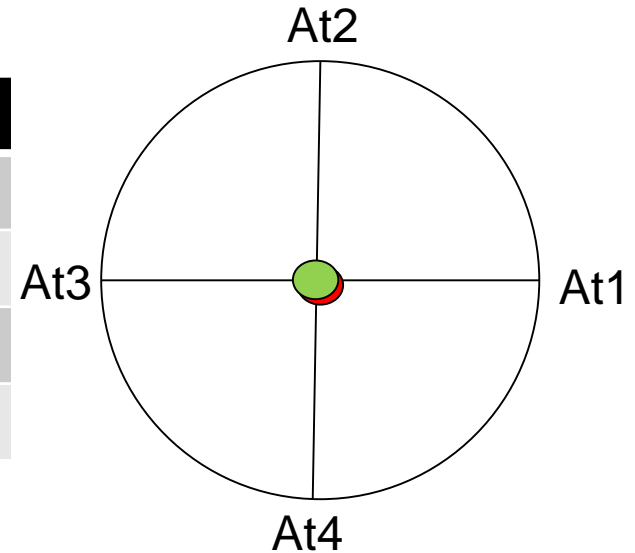
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$Ax = [1, 0, -1, 0]$$

$$Ay = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$Px = \frac{10 * 1 + 10 * 0 + 5 * (-1) + 5 * 0}{10 + 10 + 5 + 5} = \frac{10 - 5}{30} = 0,16$$

$$Py = \frac{10 * 0 + 10 * 1 + 5 * 0 + 5 * (-1)}{10 + 10 + 5 + 5} = \frac{10 - 5}{30} = 0,16$$

Observação: o círculo verde foi deslocado para vermos a sobreposição, mas na prática não veríamos o círculo vermelho

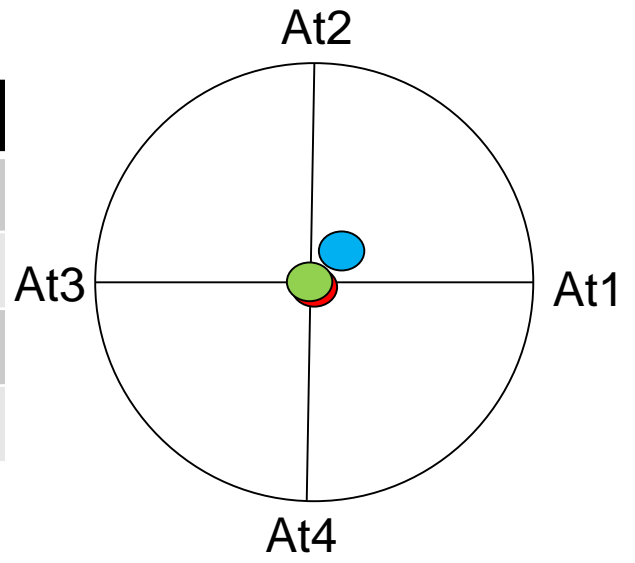
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$A_x = [1, 0, -1, 0]$$

$$A_y = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 10 * 0 + 5 * (-1) + 5 * 0}{10 + 10 + 5 + 5} = \frac{10 - 5}{30} = 0,16$$

$$P_y = \frac{10 * 0 + 10 * 1 + 5 * 0 + 5 * (-1)}{10 + 10 + 5 + 5} = \frac{10 - 5}{30} = 0,16$$

Observação: o círculo verde foi deslocado para vermos a sobreposição, mas na prática não veríamos o círculo vermelho

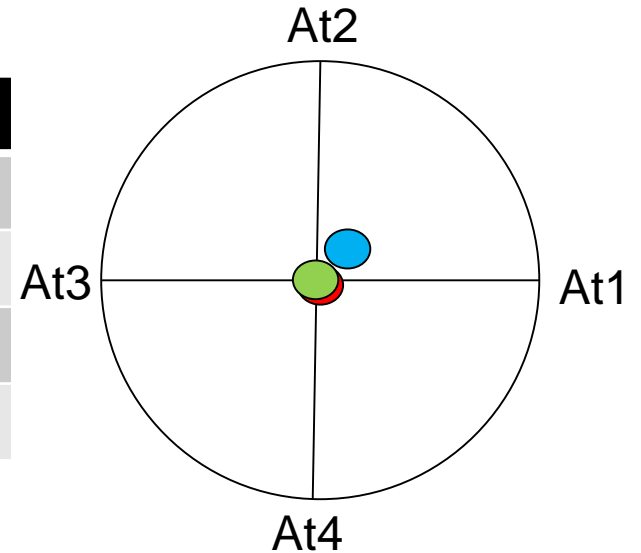
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$A_x = [1, 0, -1, 0]$$

$$A_y = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0



$$P_x = \frac{10 * 1 + 0 * 0 + 5 * (-1) + 0 * 0}{10 + 0 + 5 + 0} = \frac{10 - 5}{15} = 0,33$$

$$P_y = \frac{10 * 0 + 0 * 1 + 5 * 0 + 0 * (-1)}{10 + 0 + 5 + 0} = \frac{0 - 0}{15} = 0$$

Observação: o círculo verde foi deslocado para vermos a sobreposição, mas na prática não veríamos o círculo vermelho

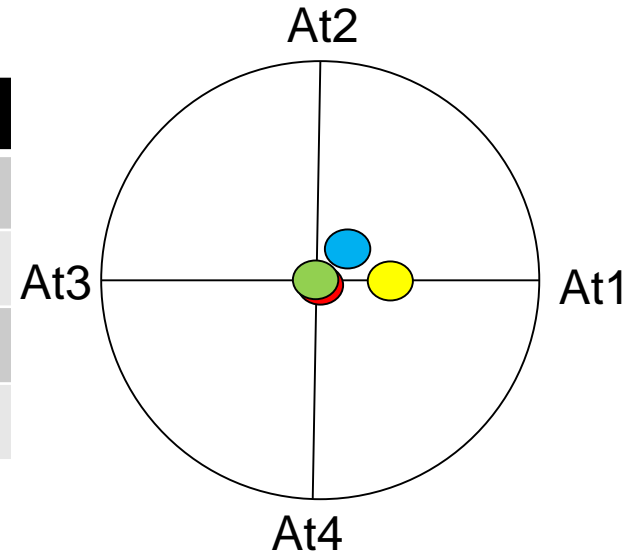
RadViz

$$p = \frac{\sum_{j=0}^{N-1} (d_j A_j)}{\sum_{j=0}^{N-1} d_j}$$

$$Ax = [1, 0, -1, 0]$$

$$Ay = [0, 1, 0, -1]$$

Atributo 1	Atributo 2	Atributo 3	Atributo 4
10	10	10	10
10	0	10	0
10	10	5	5
10	0	5	0

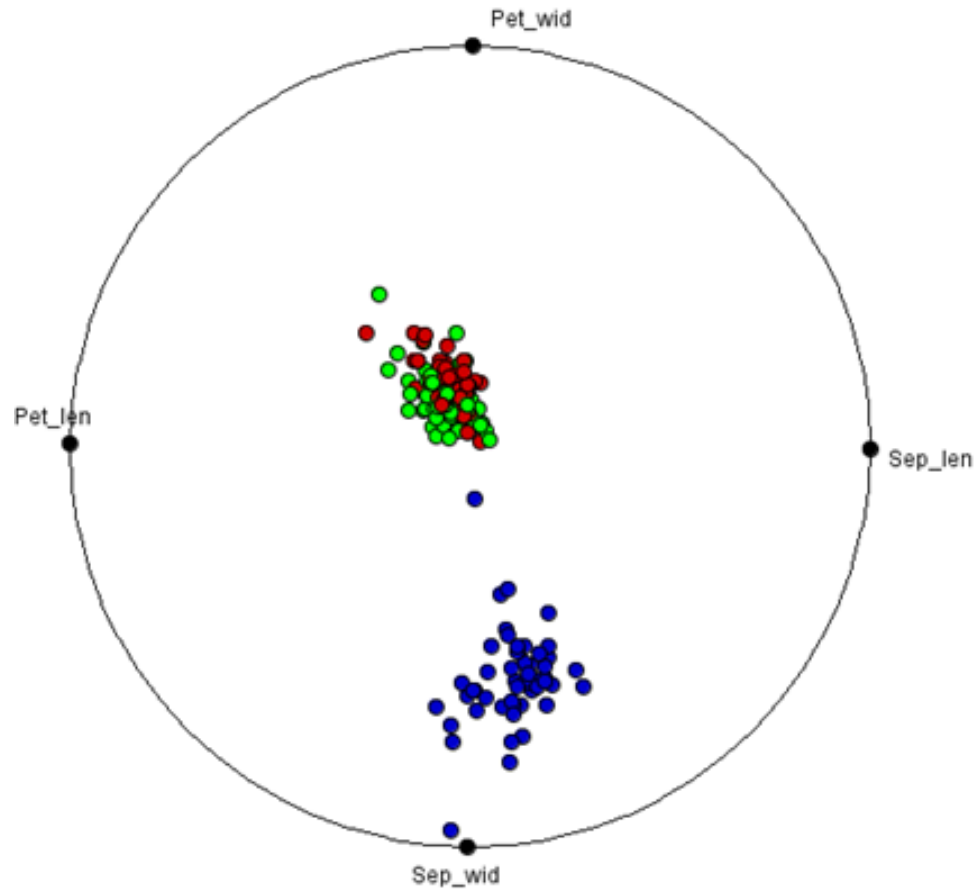


$$Px = \frac{10 * 1 + 0 * 0 + 5 * (-1) + 0 * 0}{10 + 0 + 5 + 0} = \frac{10 - 5}{15} = 0,33$$

$$Py = \frac{10 * 0 + 0 * 1 + 5 * 0 + 0 * (-1)}{10 + 0 + 5 + 0} = \frac{0 - 0}{15} = 0$$

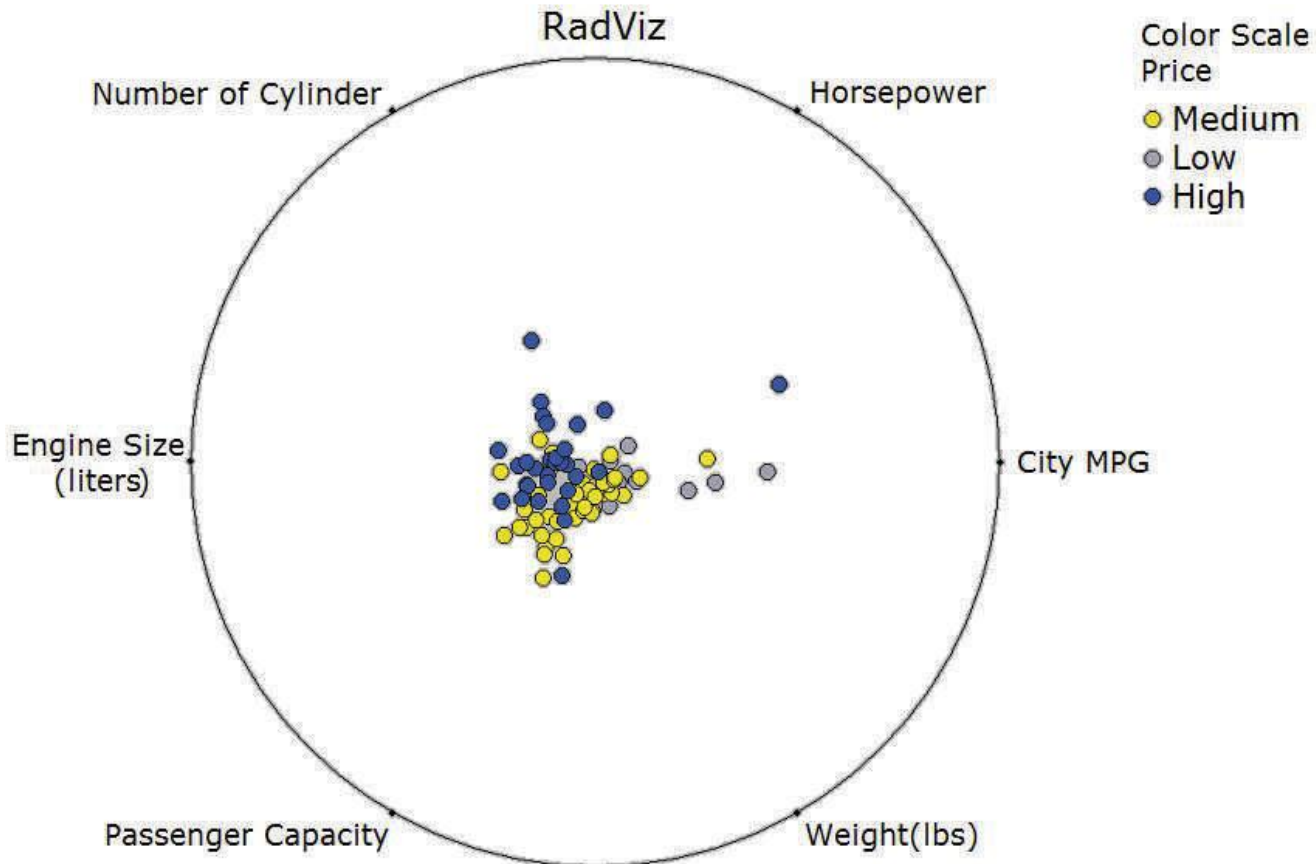
RadViz

- Exemplo para conjunto Iris



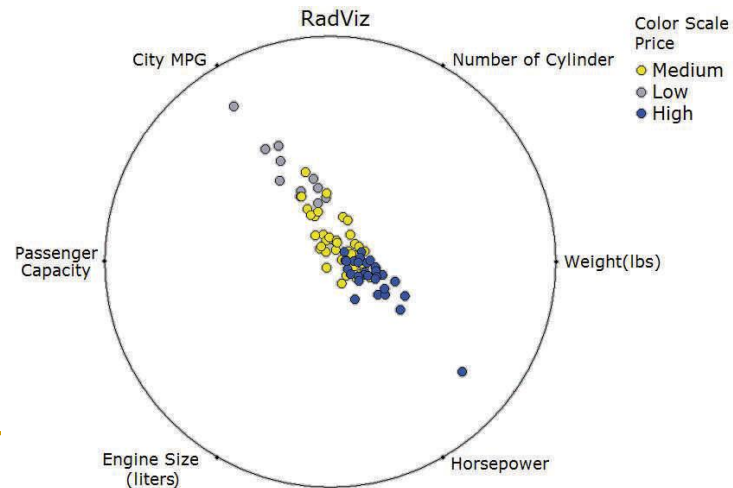
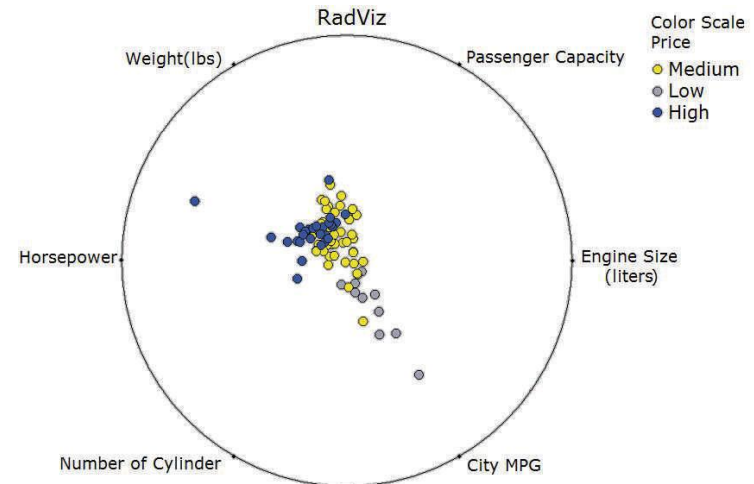
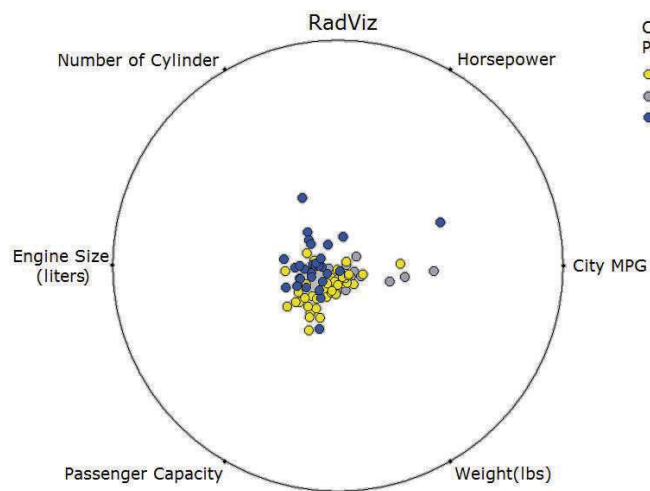
RadViz

■ Exemplo para conjunto de carros



RadViz

■ Exemplo para conjunto de carros

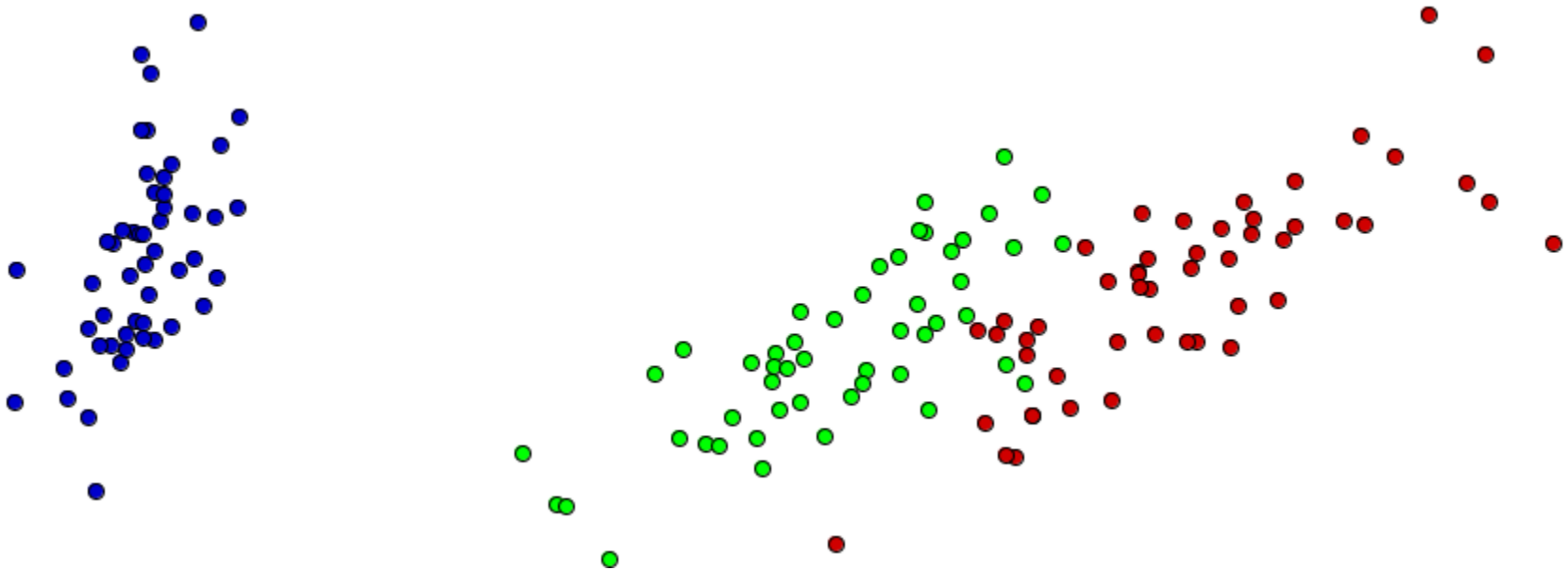


Projeção Multidimensional

- Técnicas de Projeção Multidimensional são utilizadas como uma maneira de manipular o crescimento das dimensões dos conjuntos de dados
- Essa técnica auxilia a identificar estruturas e relacionamentos nos conjuntos de dados original
 - Procuram manter as relações presentes no espaço multidimensional no espaço projetado

Projeção Multidimensional

- Projeção Multidimensional do conjunto Iris



Projeção Multidimensional

- Multidimensional Scaling (MDS) é uma importante classe de algoritmos de projeção multidimensional utilizada em visualização de informação
- Existem diferentes variações e algoritmos de MDS

Projeção Multidimensional

- A ideia básica de uma técnica de MDS proposta por Kruskal (1978) segue os seguintes passos
 1. Dada uma Matriz com N instâncias e M dimensões, cria-se uma matriz D_m de dimensão N x N com a similaridade entre cada par de instâncias
 - A distância Euclidiana pode ser utilizada como medida

DADOS

ID	At1	At2	At3	At4
1	5	3	1	0
2	4	3	9	0
3	2	3	5	2
4	8	2	2	9
5	8	2	2	7



D_m

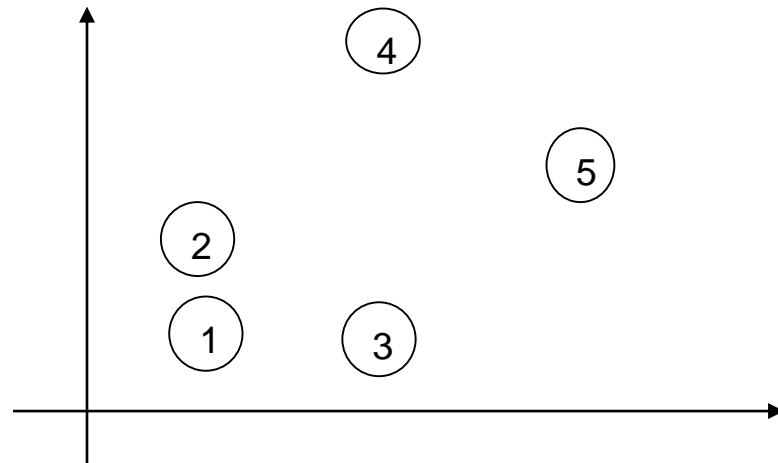
	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Projeção Multidimensional

- A ideia básica de uma técnica de MDS proposta por Kruskal (1978) segue os seguintes passos
 2. Assumindo que os dados são projetados em um espaço de dimensão K , cria-se uma matriz L de dimensão $N \times K$ com as posições das instâncias no espaço projetado
 1. Podem ser aleatoriamente espalhados no espaço K ou utilizar outra técnica para isso (e.g., PCA, Fastmap)

L

ID	At1	At2
1	1	1
2	1	2
3	2	1
4	2	4
5	4	4



Projeção Multidimensional

- A ideia básica de uma técnica de MDS proposta por Kruskal (1978) segue os seguintes passos
 3. Calcule uma matriz D_k de dimensão $N \times N$ com as similaridades entre os pares de pontos de L (espaço projetado)

L

ID	At1	At2
1	1	1
2	1	2
3	2	1
4	2	4
5	4	4



D_k

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0

Projeção Multidimensional

- A ideia básica de uma técnica de MDS proposta por Kruskal (1978) segue os seguintes passos
 4. Calcule o valor do stress (S) medido a partir das diferenças entre D_m e D_k
 - Exemplos de função de stress

$$S = \sum_{ij} (d_{ij} - \bar{d}_{ij})^2$$

$$S = \sum_{ij} \frac{(d_{ij} - \bar{d}_{ij})^2}{\sum_{ij} d_{ij}^2}$$

D_m

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

D_k

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0

Projeção Multidimensional

- A ideia básica de uma técnica de MDS proposto por Kruskal (1978) segue os seguintes passos
 3. Calcule uma matriz D_k de dimensão $N \times N$ com as similaridades entre os pares de pontos de L (espaço projetado)
 4. Calcule o valor do stress (S) medido a partir das diferenças entre D_m e D_k
 - Exemplos de função de stress

$$S = \sum_{ij} (d_{ij} - \bar{d}_{ij})^2$$

$$S = \sum_{ij} \frac{(d_{ij} - \bar{d}_{ij})^2}{\sum_{ij} d_{ij}^2}$$

5. Se o stress S é suficientemente pequeno ou nenhuma mudança significativa ocorreu em iterações recentes, o algoritmo termina
6. Caso contrário, tenta movimentar as instâncias projetadas (matriz L) na direção que diminuirá os níveis de stress
7. Retorne ao passo 3

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 1 <--> 2

Dm = 8,0

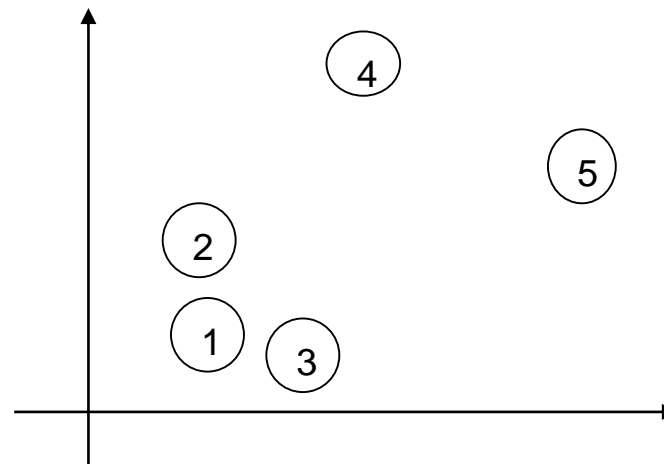
Dk = 1,0

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0

L

ID	At1	At2
1	1	1
2	1	2
3	2	1
4	2	4
5	4	4



Atenção! As distâncias do desenho são ilustrativas

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

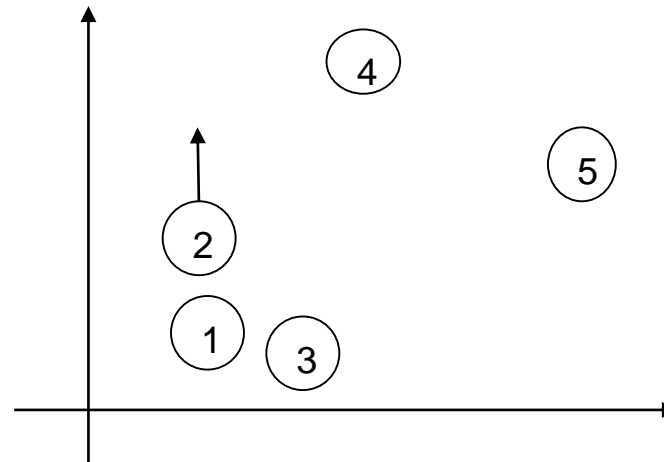
Comparando 1 <--> 2

$D_m = 8,0$

$D_k = 1,0$

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



L

ID	At1	At2
1	1	1
2	1	2
3	2	1
4	2	4
5	4	4

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 1 <--> 2

$D_m = 8,0$

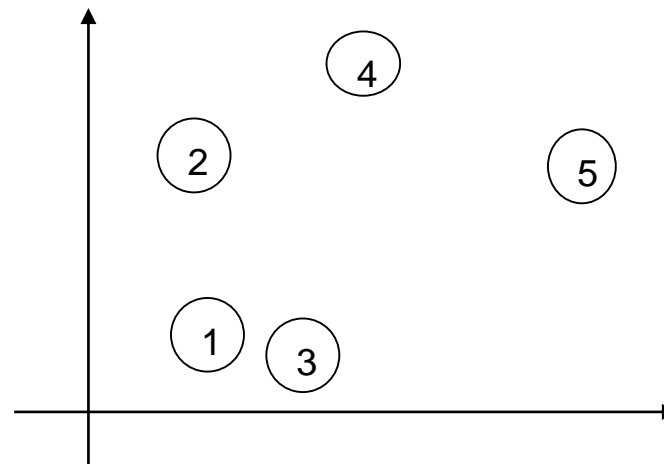
$D_k = 1,0$

L

ID	At1	At2
1	1	1
2	2	2
3	2	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 1 <--> 3

$D_m = 8,0$

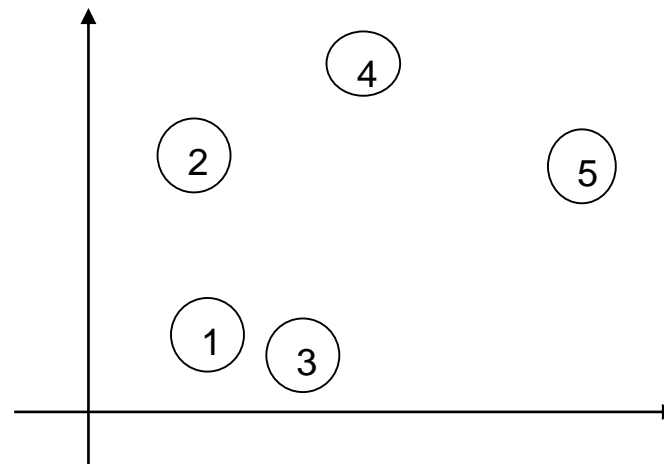
$D_k = 1,0$

L

ID	At1	At2
1	1	1
2	2	2
3	2	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 1 <--> 3

$D_m = 8,0$

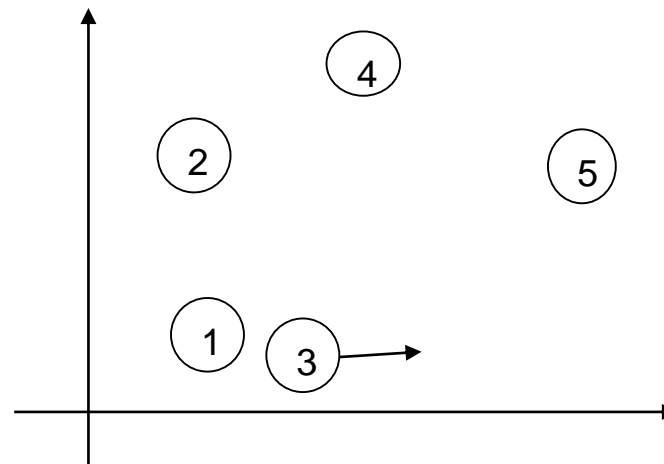
$D_k = 1,0$

L

ID	At1	At2
1	1	1
2	2	2
3	2	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 1 <--> 3

$D_m = 8,0$

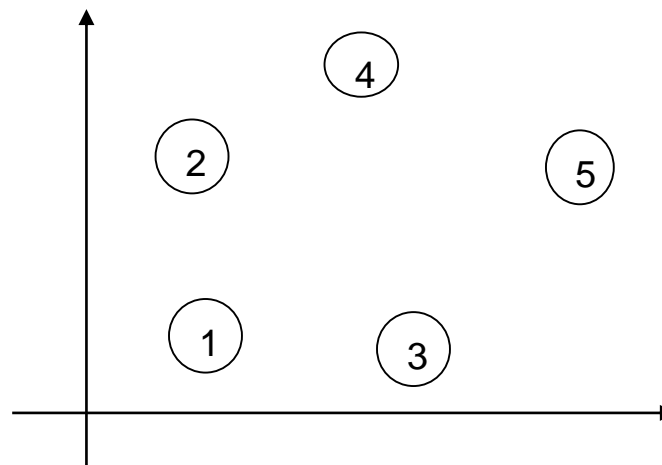
$D_k = 1,0$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

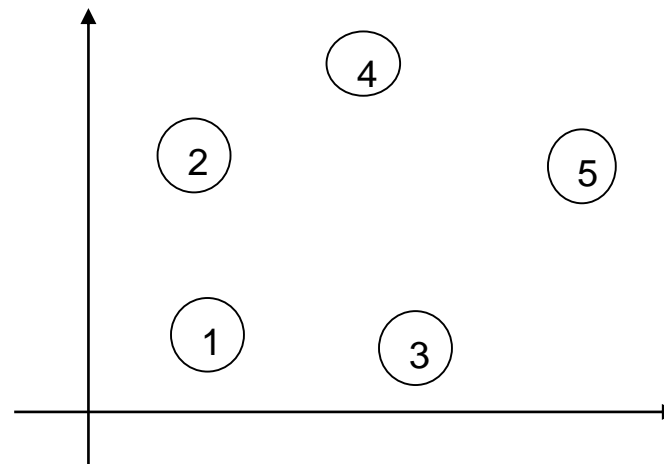
Comparando 1 <--> 4
 $D_m = 9,5$
 $D_k = 3,1$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

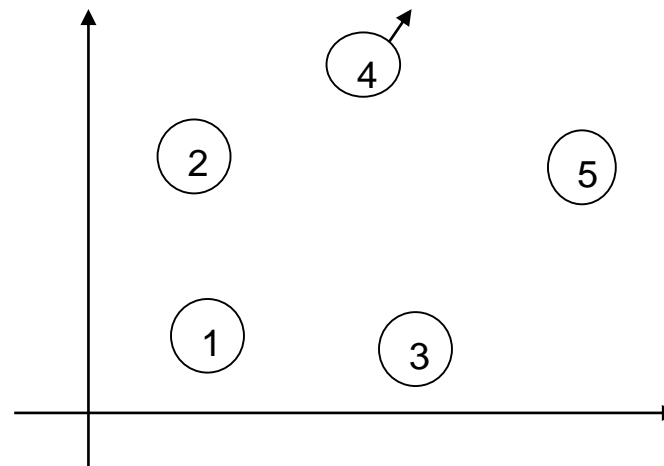
Comparando 1 <--> 4
 $D_m = 9,5$
 $D_k = 3,1$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

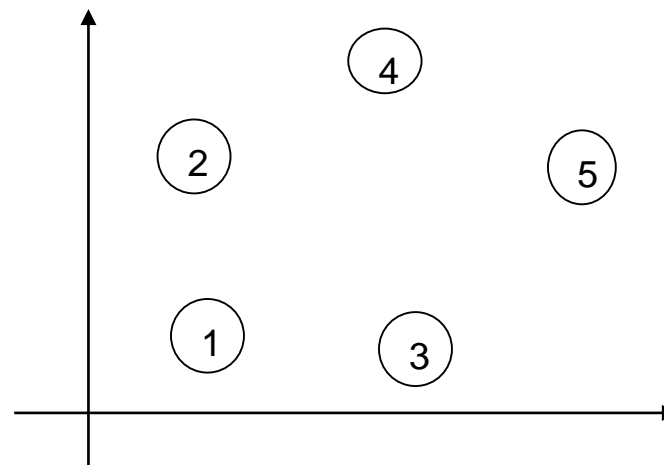
Comparando 1 <--> 4
 $D_m = 9,5$
 $D_k = 3,1$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

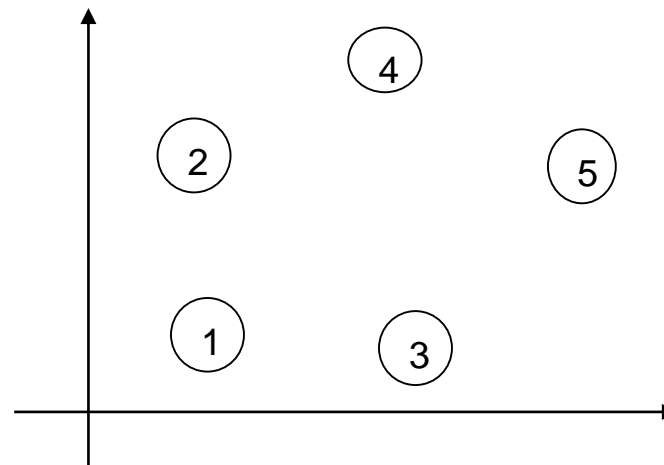
Comparando 1 <--> 5
 $D_m = 7,7$
 $D_k = 4,2$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

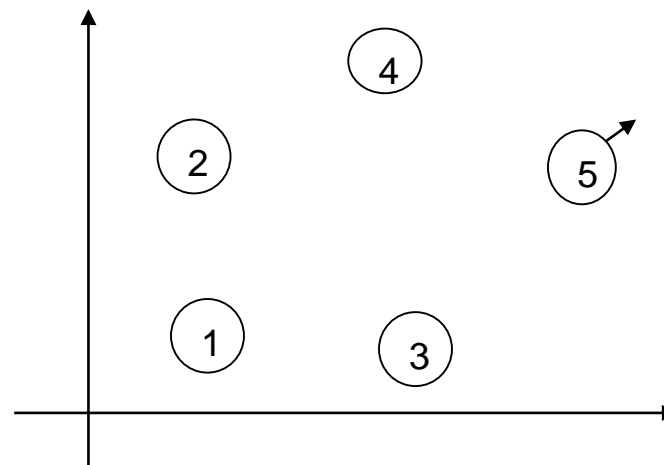
Comparando 1 <--> 5
 $D_m = 7,7$
 $D_k = 4,2$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4	4

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

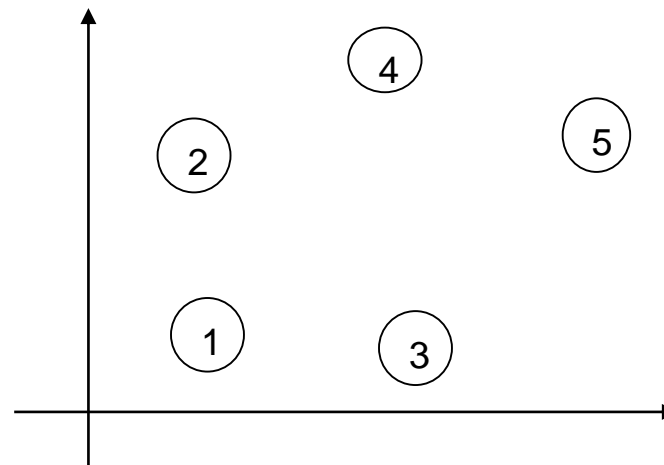
Comparando 1 <--> 5
 $D_m = 7,7$
 $D_k = 4,2$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4,2	4,2

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,0
5	4,2	3,6	3,6	2,0	0,0



Não estou atualizando Dk

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 4 <--> 5

$D_m = 2,0$

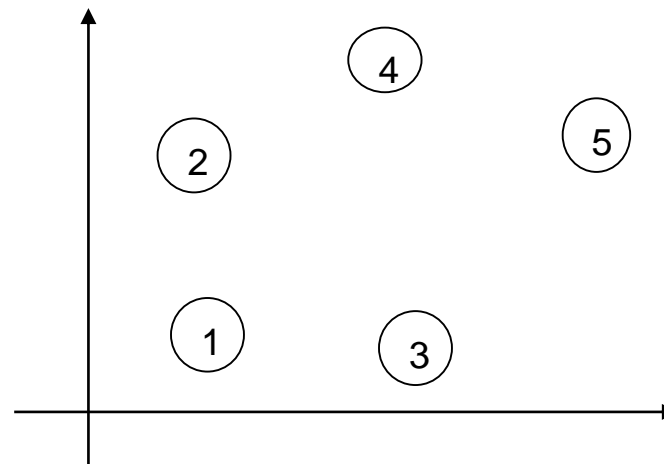
$D_k = 2,3$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4,2	4,2

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,3
5	4,2	3,6	3,6	2,3	0,0



Atualizei Dk para 4 e 5

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 4 <--> 5

$D_m = 2,0$

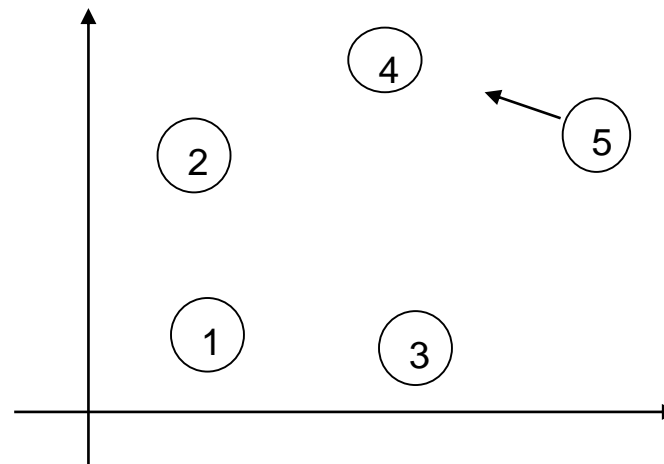
$D_k = 2,3$

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	4,2	4,2

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	2,3
5	4,2	3,6	3,6	2,3	0,0



Atualizei Dk para 4 e 5

Projeção Multidimensional

- MDS proposto por Kruskal (1978)

Dm

	1	2	3	4	5
1	0,0	8,0	8,0	9,5	7,7
2	8,0	0,0	4,5	12,1	10,7
3	8,0	4,5	0,0	10,4	9,0
4	9,5	12,1	10,4	0,0	2,0
5	7,7	10,7	9,0	2,0	0,0

Comparando 4 <--> 5

Dm = 2,0

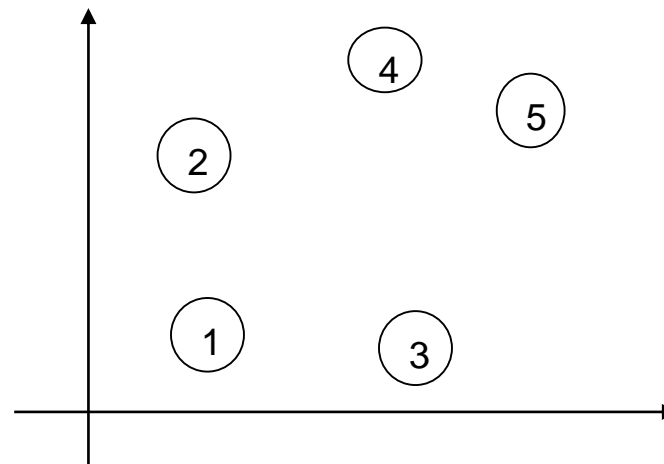
Dk = 2,3

L

ID	At1	At2
1	1	1
2	2	2
3	3	1
4	2	4,5
5	3,7	4,3

Dk

	1	2	3	4	5
1	0,0	1,0	1,0	3,1	4,2
2	1,0	0,0	1,4	2,2	3,6
3	1,0	1,4	0,0	3,0	3,6
4	3,1	2,0	3,0	0,0	1,7
5	4,2	3,6	3,6	1,7	0,0



Atualizei Dk para 4 e 5

Projeção Multidimensional

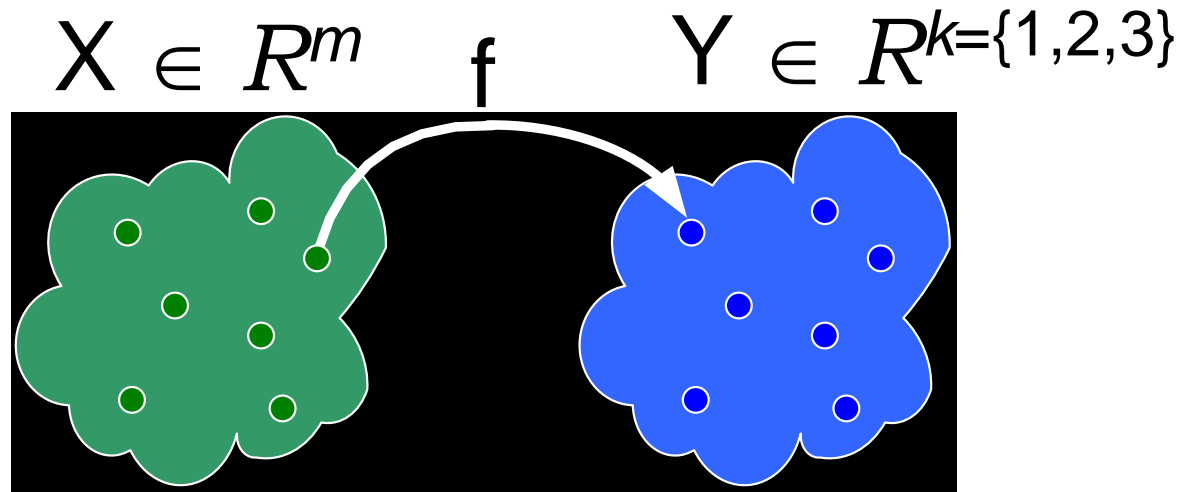
- A ideia básica de uma técnica de MDS proposto por Kruskal (1978) segue os seguintes passos
 3. Calcule uma matriz D_k de dimensão $N \times N$ com as similaridades entre os pares de pontos de L (espaço projetado)
 4. Calcule o valor do stress (S) medido a partir das diferenças entre D_m e D_k
 - Exemplos de função de stress

$$S = \sum_{ij} (d_{ij} - \bar{d}_{ij})^2$$

$$S = \sum_{ij} \frac{(d_{ij} - \bar{d}_{ij})^2}{\sum_{ij} d_{ij}^2}$$

5. Se o stress S é suficientemente pequeno ou nenhuma mudança significativa ocorreu em iterações recentes, o algoritmo termina
6. Caso contrário, tenta movimentar as instâncias projetadas (matriz L) na direção que diminuirá os níveis de stress
7. Retorne ao passo 3

Projeção Multidimensional



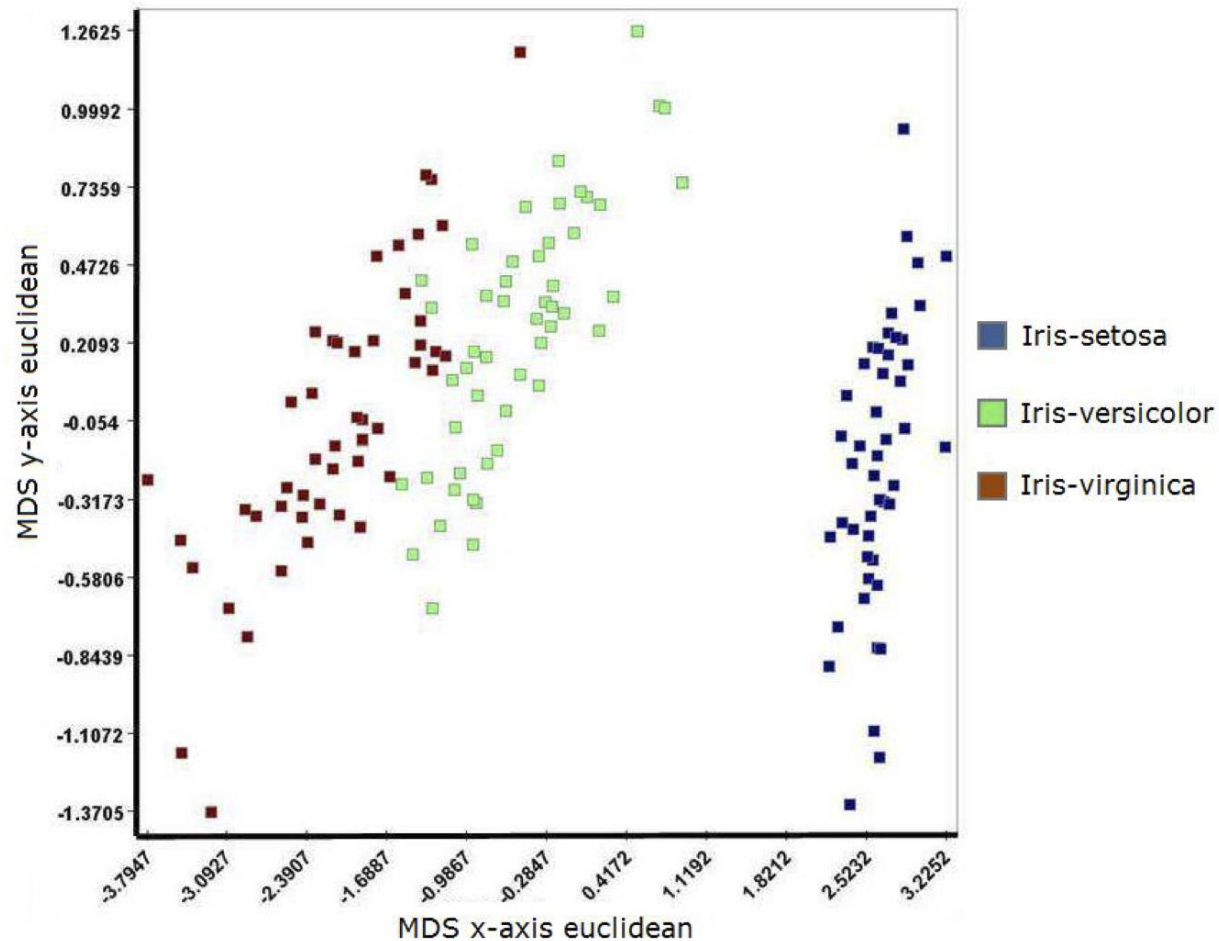
$$\delta: x_i, x_j \rightarrow \mathbb{R}, x_i, x_j \in X$$

$$d: y_i, y_j \rightarrow \mathbb{R}, y_i, y_j \in Y$$

$$f: X \rightarrow Y, |\delta(x_i, x_j) - d(f(x_i), f(x_j))| \approx 0, \forall x_i, x_j \in X$$

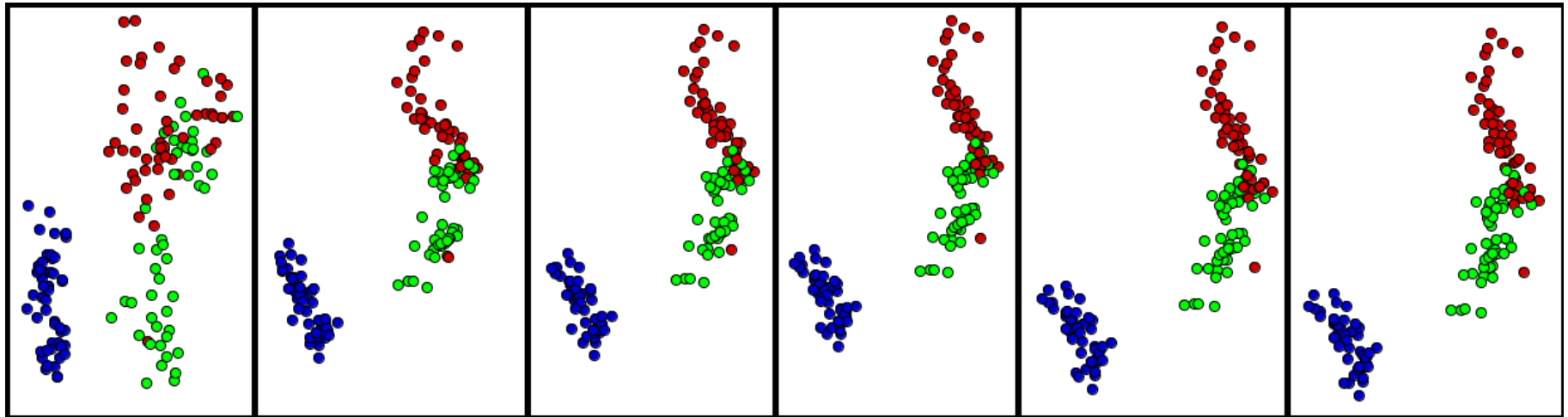
Projeção Multidimensional

■ Projeção MDS do conjunto Iris



Projeção Multidimensional

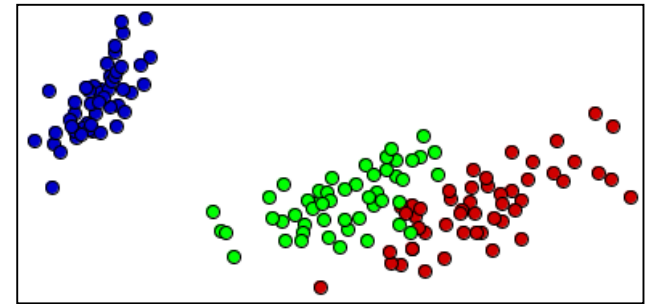
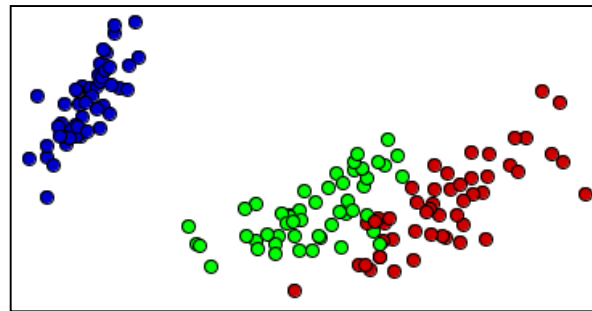
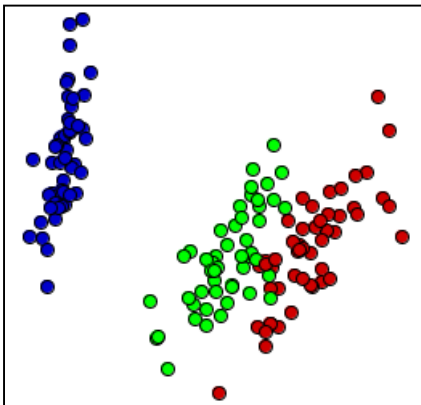
- Exemplo de iterações de uma técnica de MDS
 - Os valores de stress são, respectivamente,
 - 0,17; 0,016; 0,009; 0,006; 0,005; 0,005



Inicialmente projetado com Nearest Neighbor Projection (NNP)

Projeção Multidimensional

- Exemplo de iterações de uma técnica de MDS
 - Os valores de stress são, respectivamente,
 - 0,03; 0,001; 0,001



Inicialmente projetado com Fastmap

Referências

- Ward, M., Grinstein, G. G., Keim, D.
 - Interactive data visualization foundations, techniques, and applications. Natick, Mass., A K Peters, 2010.
 - Capítulos 8 e 9
- Daniel A. Keim
 - Designing Pixel-Oriented Visualization Techniques: Theory and Applications. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 6, NO. 1, 2000
- Daniel A. Keim and Hans-Peter Kriegel
 - VisDB: Database Exploration Using Multidimensional Visualization. IEEE Computer Graphics and Applications, vol. 14, no. 5, pp. 40-49, Sept. 1994

Referências

- Maria Cristina Ferreira de Oliveira e Haim Levkowitz
 - From visual data exploration to visual data mining: a survey. IEEE Transactions on Visualization and Computer Graphics 9 (3), 378-394, 2003
- Tese de Doutorado
 - Milton Hirokazu Shimabukuro, Visualizações Temporais em uma Plataforma de Software Extensível e Adaptável, ICMC/USP, 2004
- Ronald M Pickett, Georges G. Grinstein
 - Iconographic Displays for Visualizing Multidimensional Data, Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Beijing and Shenyang, China, 514-419, 1998

Referências

- Alfred Inselberg
 - “The Plane with Parallel Coordinates.” *The Visual Computer* 1:2 (1985), 69–91
- Patrick Hoffman, Georges Grinstein, Kenneth Marx, Ivo Grosse, and Eugene Stanley
 - “DNA Visual and Analytic Data Mining.” In *VIS '97: Proceedings of the 8th Conference on Visualization '97*, pp. 437–ff. Los Alamitos, CA: IEEE Computer Society Press, 1997.
- Brian Johnson and Ben Shneiderman
 - “Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures.” In *VIS '91: Proceedings of the 2nd Conference on Visualization '91*, pp. 284–291. Los Alamitos, CA: IEEE Computer Society Press, 1991.

Referências

- John Stasko and Eugene Zhang.
 - “Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations.” In Proceedings of the IEEE Symposium on Information Visualization, pp. 57–65. Los Alamitos, CA: IEEE Computer Society, 2000.
- George G. Robertson, Jock D. Mackinlay, and Stuart K. Card.
 - “Cone Trees: Animated 3D Visualizations of Hierarchical Information.” In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 189–194. New York: ACM Press, 1991.

Referências

- A. Astel, K. Astel, M. Biziuk, J. Namieśnik.
 - Clasification of Drinking Water Samples Using the Chernoff's Faces Visualization Approach, Polish J. of Environ. Stud. Vol. 15, No. 5 (2006), 691-697
- Shneiderman
 - Bem Shneiderman. The eyes have it: a task by data type taxonomy for information visualization. In Proceedings of the 1996 IEEE Symposium on Visual Languages, pp. 336-343. Washington, DC: IEEE Computer Society, 1996
- Keim
 - Daniel A. Keim. Information Visualization and Visual Data Mining. IEEE Transactions on Visualization and Computer Graphics, 8:1 (2002), 1-8

Referências

- J. B. Kruskal and M. Wish
 - Multidimensional Scaling. Quantitative Applications in the Social Sciences Series, Newbury Park: Sage Publications, 1978.
- História da Treemap
 - <http://www.cs.umd.edu/hcil/treemap-history/>