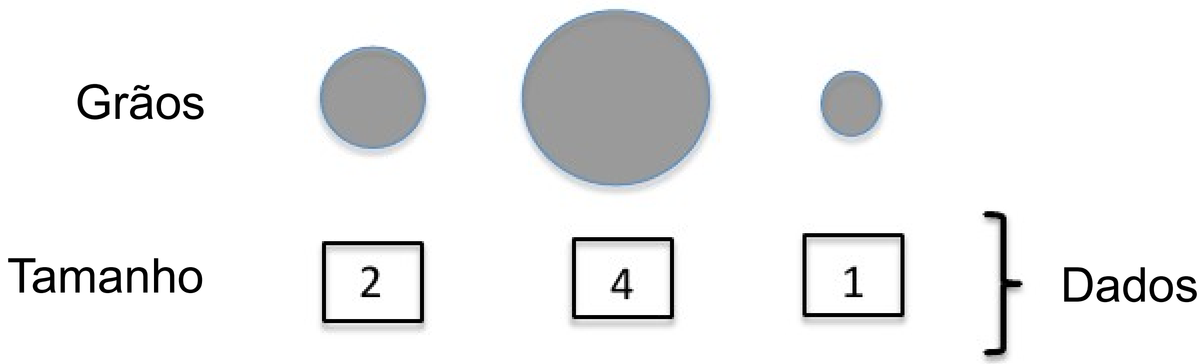
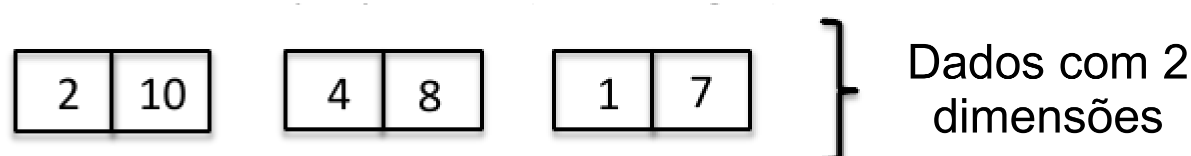
De formigas mortas para dados

- Coletar grãos de tamanhos semelhantes



- Grãos com duas propriedades: (tamanho, peso)

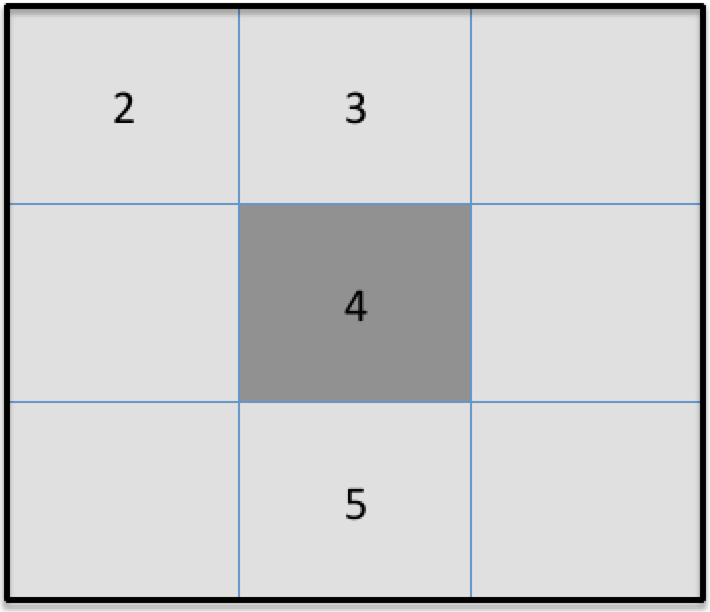


Agrupando um conjunto de dados

- Coletando dados similares:
  - Se a formiga ainda não está carregando um dado e se depara com um, pegue-o se observar poucos dados similares na sua área de visão
  - Se a formiga está carregando um dado, deixe-o se observar muitos outros dados similares na sua área de visão

- No caso das formigas mortas, simplesmente conta-se o número de formigas mortas e calcula-se a fração em relação sua área de visão.
  - Todas elas possuem uma única propriedade: estão todas mortas
- No caso de 'grãos' calcula-se a similaridade entre os grãos que se deseja pegar ou largar na sua área de visão
- Diferencia-se entre os dois pontos de vista através de uma função *f*.
  - Na primeira visão, *f* é somente a fração de formigas mortas percebida.
  - Na segunda visão, *f* é uma similaridade entre os grãos.

Distância entre um par de dados

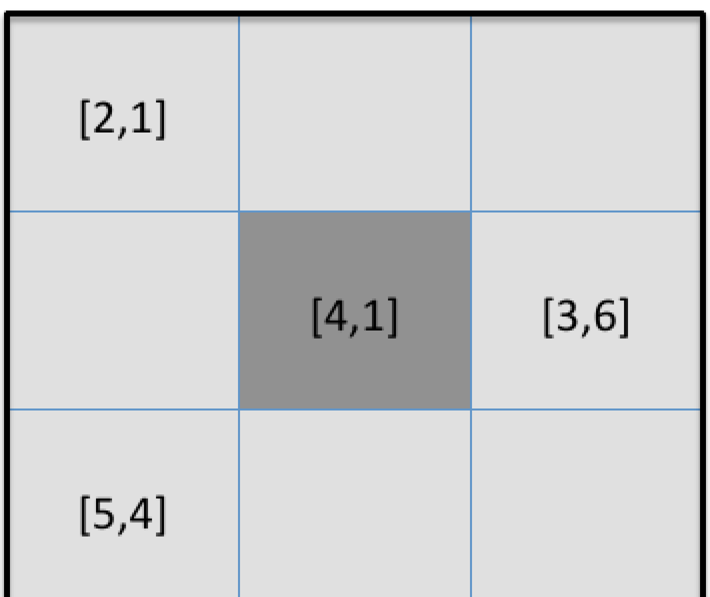


Distância Euclidiana

$$D(4,2) = \sqrt{(4-2)^2} = 2$$

$$D(4,3) = \sqrt{(4-3)^2} = 1$$

$$D(4,5) = \sqrt{(4-5)^2} = 1$$



$$D([4,1],[2,1]) = \sqrt{(4-2)^2 + (1-1)^2} = 2$$

$$D([4,1],[3,6]) = \sqrt{(4-3)^2 + (1-6)^2} = 5.1$$

$$D([4,1],[5,4]) = \sqrt{(4-5)^2 + (1-4)^2} = 3.16$$

Distância entre um par de dados

- Distância Euclidiana

Vetor Xi

$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	...	$x_{i,n}$

Vetor Xj

$x_{j,1}$	$x_{j,2}$	$x_{j,3}$	...	$x_{j,n}$

$$D(x_i, x_j) = \sqrt{\sum_{k=1}^{n} \left(x_{i,k} - x_{j,k}\right)^2}$$

Similaridade entre um dado e sua vizinhança

- Similaridade da vizinhança ou medida de densidade

$$f(i) = \begin{cases} \frac{1}{s^2} \sum_j (1 - d(i, j) / \alpha) & \text{if} \quad f(i) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Quanto menor as distâncias tendem a ser, maior a similaridade tende a ser com a dado que pretende-se carregar ou largar (x_i) em relação às células da vizinhança (*s* x *s*)
- $f(x_i)$ deve variar no intervalo [0, 1]
- *s* representa a quantidade de dados na vizinhança
- *α* é o fator que define a escala para a dissimilaridade. Não deve ser nem muito maior nem muito menor que o valor médio de distância esperado (definido empiricamente)

Pegando e Largando um Dado

$$p_p(x_i) = \left(\frac{k_1}{k_1 + f(x_i)} \right)^2 \qquad p_d(x_i) = \left(\frac{f(x_i)}{k_2 + f(x_i)} \right)^2$$

- Como o menor valor que $f(x_i)$ pode assumir é 0, k1 e k2 devem ser maiores que 0
- Como o maior valor que $f(x_i)$ pode assumir é 1, k1 e k2 não devem ser superestimados de maneira a suprimir o valor de $f(x_i)$
- Note que Pp varia de k1/(k1 + 1) até 1 e
- Pd varia de 0 até 1/(k2 + 1)
- k1 e k2 variam entre [0, 1]

Pseudo Código

- Crie uma grid NxM
- Defina o raio de visão das formigas
- Distribua aleatoriamente os dados pela grid
- Distribua aleatoriamente Q formigas na grid, com não mais de uma formiga por célula
- Repita os passo por um número finito de iterações
  - Para cada formiga na grid, em sua respectivas células:
    - Se a formiga não está carregando um dado e sua célula contém um dado xi então
      - Calcule f(xi)
      - Pegue o dado com probabilidade Pp(xi)
    - Se a formiga está carregando um dado xi e sua célula está livre, então
      - Calcule f(xi)
      - Solte o dado com probabilidade Pd(xi)
    - Desloque a formiga para um célula vizinha não ocupada por outra formiga