Aula 18 - Algoritmos de Agrupamento

João Florindo

Instituto de Matemática, Estatística e Computação Científica
Universidade Estadual de Campinas - Brasil
florindo@unicamp.br

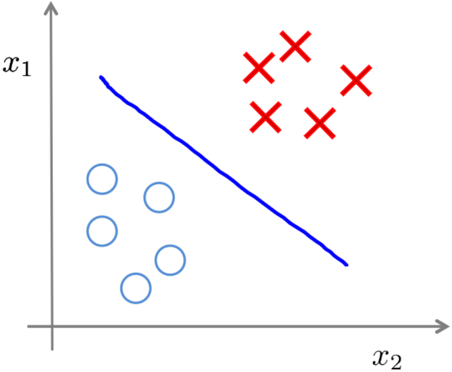
Outline

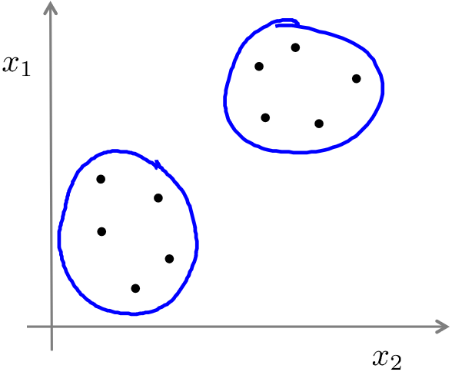
- Antes:

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \cdots, (x^{(m)}, y^{(m)})\}.$$

- Agora:

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \cdots, x^{(m)}\}.$$





Aplicações

- Segmentação de mercado
- Redes sociais
- Clusters de computadores
- Astronomia

Aplicações

- Segmentação de mercado
- Redes sociais
- Clusters de computadores
- Astronomia

Aplicações

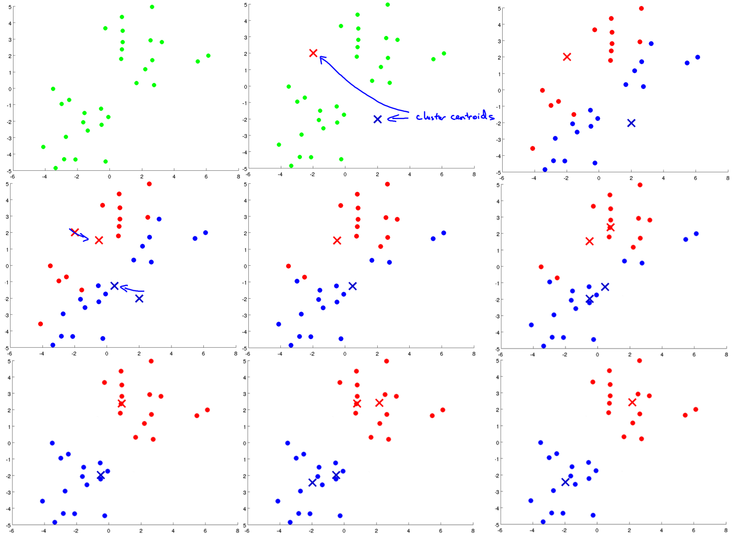
- Segmentação de mercado
- Redes sociais
- Clusters de computadores
- Astronomia

Aplicações

- Segmentação de mercado
- Redes sociais
- Clusters de computadores
- Astronomia

Outline

K-means



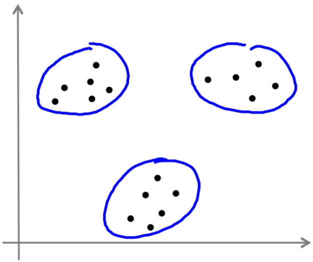
K-means

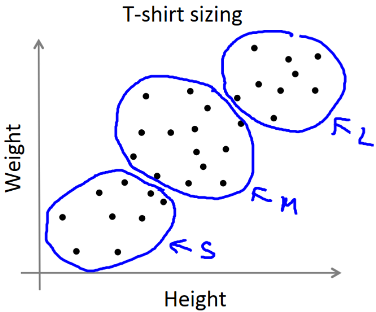
Entradas:

- K (número de clusters)
- $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$ (convenção $x_0 = 1$)

Inicializar aleatoriamente K centroides $\mu_1, \mu_2, \cdots, \mu_K \in \mathbb{R}^n$
Repita {
    Para $i = 1$ até m *% passo de atribuição do cluster*
        $c^{(i)} := \text{argmin}_k \|x^{(i)} - \mu_k\|^2$ *% índice (de 1 a K) do*
                % centroide mais próximo de $x^{(i)}$
    Para $k = 1$ até K *% movimento do centroide*
        $\mu_k :=$ média dos pontos atribuídos ao cluster k
}

K-means





Função Objetivo

Notação

- $c^{(i)}$: índice do cluster $(1, 2, \cdots, K)$ ao qual o exemplo $x^{(i)}$ está atribuído
- μ_k: centroide k ($\mu_k \in \mathbb{R}^n$)
- $\mu_{c^{(i)}}$: centroide do cluster ao qual o exemplo $x^{(i)}$ foi atribuído

Função objetivo:

$$J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K) = \frac{1}{m} \sum_{i=1}^{m} \left\| x^{(i)} - \mu_{c^{(i)}} \right\|^2$$

Otimização:

$$\underset{\substack{c^{(1)}, \cdots, c^{(m)} \\ \mu_1, \cdots, \mu_K}}{\operatorname{argmin}} \; J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K).$$

A função J é chamada de **distorção**.

Função Objetivo

- O algoritmo é de coordenada descendente sobre J.
- o primeiro laço do algoritmo (atribuição do cluster) minimiza J w.r.t. $c^{(1)}, \cdots, c^{(m)}$, mantendo-se $\mu_1, \cdots, \mu_K$ fixos.
- Já o segundo laço (movimento do centroide) minimiza J w.r.t. $\mu_1, \cdots, \mu_K$.
- J é não-convexa (sem garantia de convergência para mínimo global).

Função Objetivo

- O algoritmo é de coordenada descendente sobre J.
- o primeiro laço do algoritmo (atribuição do cluster) minimiza J w.r.t. $c^{(1)}, \cdots, c^{(m)}$, mantendo-se $\mu_1, \cdots, \mu_K$ fixos.
- Já o segundo laço (movimento do centroide) minimiza J w.r.t. $\mu_1, \cdots, \mu_K$.
- J é não-convexa (sem garantia de convergência para mínimo global).

Função Objetivo

- O algoritmo é de coordenada descendente sobre J.
- o primeiro laço do algoritmo (atribuição do cluster) minimiza J w.r.t. $c^{(1)}, \cdots, c^{(m)}$, mantendo-se $\mu_1, \cdots, \mu_K$ fixos.
- Já o segundo laço (movimento do centroide) minimiza J w.r.t. $\mu_1, \cdots, \mu_K$.
- J é não-convexa (sem garantia de convergência para mínimo global).

Função Objetivo

- O algoritmo é de coordenada descendente sobre J.
- o primeiro laço do algoritmo (atribuição do cluster) minimiza J w.r.t. $c^{(1)}, \cdots, c^{(m)}$, mantendo-se $\mu_1, \cdots, \mu_K$ fixos.
- Já o segundo laço (movimento do centroide) minimiza J w.r.t. $\mu_1, \cdots, \mu_K$.
- J é não-convexa (sem garantia de convergência para mínimo global).

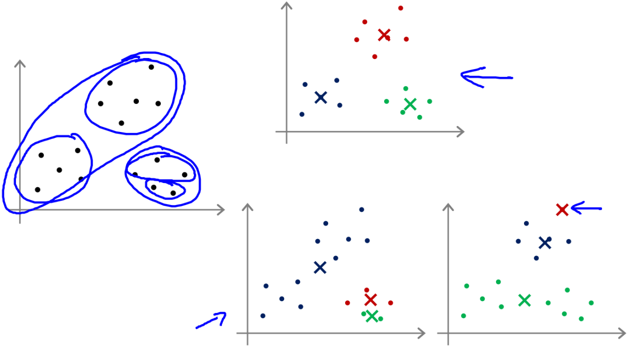
Outline

Inicialização

- Selecionar K exemplos de treinamento aleatoriamente e usar como $\mu_1, \cdots, \mu_K$.
- Diferentes inicializações geram diferentes agrupamentos (mínimos locais de J).

Inicialização

- Selecionar K exemplos de treinamento aleatoriamente e usar como $\mu_1, \cdots, \mu_K$.
- Diferentes inicializações geram diferentes agrupamentos (mínimos locais de J).



Inicialização

- Rodar o algoritmo várias vezes e tomar o caso com menor custo.

Para $i = 1$ até 100{
    Inicializar o K-means aleatoriamente.
    Rodar o K-means e obter $c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K$.
    Calcular a função de custo (distorção)
        $J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K)$
}
Tome o agrupamento com menor custo $J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K)$

- Interessante sobretudo quando K é pequeno (2 a 5, por exemplo), não é necessário para K muito grande.

Inicialização

- Rodar o algoritmo várias vezes e tomar o caso com menor custo.

Para $i = 1$ até $100\{$

Inicializar o K-means aleatoriamente.

Rodar o K-means e obter $c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K$.

Calcular a função de custo (distorção)

$J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K)$

$\}$

Tome o agrupamento com menor custo $J(c^{(1)}, \cdots, c^{(m)}, \mu_1, \cdots, \mu_K)$

- Interessante sobretudo quando K é pequeno (2 a 5, por exemplo), não é necessário para K muito grande.

Números de Clusters

- Qual o valor ideal de K?
- Método do "cotovelo" ("*elbow*").
- Plotar J em função do número de clusters e identificar o ponto a partir do qual J decresce muito mais lentamente ("cotovelo").

- PROBLEMA: decaimento pode ser homogêneo (sem "cotovelo")!

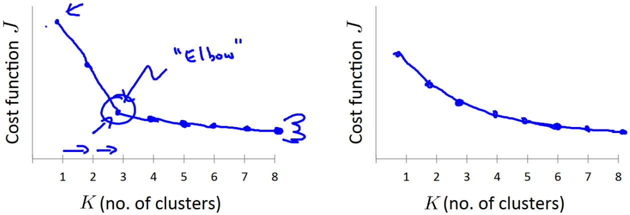
Números de Clusters

- Qual o valor ideal de K?
- Método do "cotovelo" ("*elbow*").
- Plotar J em função do número de clusters e identificar o ponto a partir do qual J decresce muito mais lentamente ("cotovelo").

- PROBLEMA: decaimento pode ser homogêneo (sem "cotovelo")!

Números de Clusters

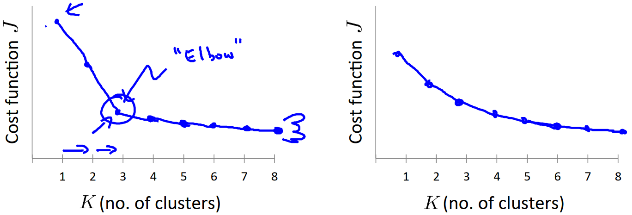
- Qual o valor ideal de K?
- Método do "cotovelo" ("*elbow*").
- Plotar J em função do número de clusters e identificar o ponto a partir do qual J decresce muito mais lentamente ("cotovelo").



- PROBLEMA: decaimento pode ser homogêneo (sem "cotovelo")!

Números de Clusters

- Qual o valor ideal de K?
- Método do "cotovelo" ("*elbow*").
- Plotar J em função do número de clusters e identificar o ponto a partir do qual J decresce muito mais lentamente ("cotovelo").



- PROBLEMA: decaimento pode ser homogêneo (sem "cotovelo")!

Números de Clusters

- Se houver uma métrica do contexto do problema que depende de K, é sempre importante levar em conta.
- NOTA: quando nenhum exemplo é atribuído a determinado centroide, o mais comum é eliminar aquele cluster.
- Outra solução, mas menos usual, é rodar novamente com outra inicialização aleatória.

Números de Clusters

- Se houver uma métrica do contexto do problema que depende de K, é sempre importante levar em conta.
- NOTA: quando nenhum exemplo é atribuído a determinado centroide, o mais comum é eliminar aquele cluster.
- Outra solução, mas menos usual, é rodar novamente com outra inicialização aleatória.

Números de Clusters

- Se houver uma métrica do contexto do problema que depende de K, é sempre importante levar em conta.
- NOTA: quando nenhum exemplo é atribuído a determinado centroide, o mais comum é eliminar aquele cluster.
- Outra solução, mas menos usual, é rodar novamente com outra inicialização aleatória.