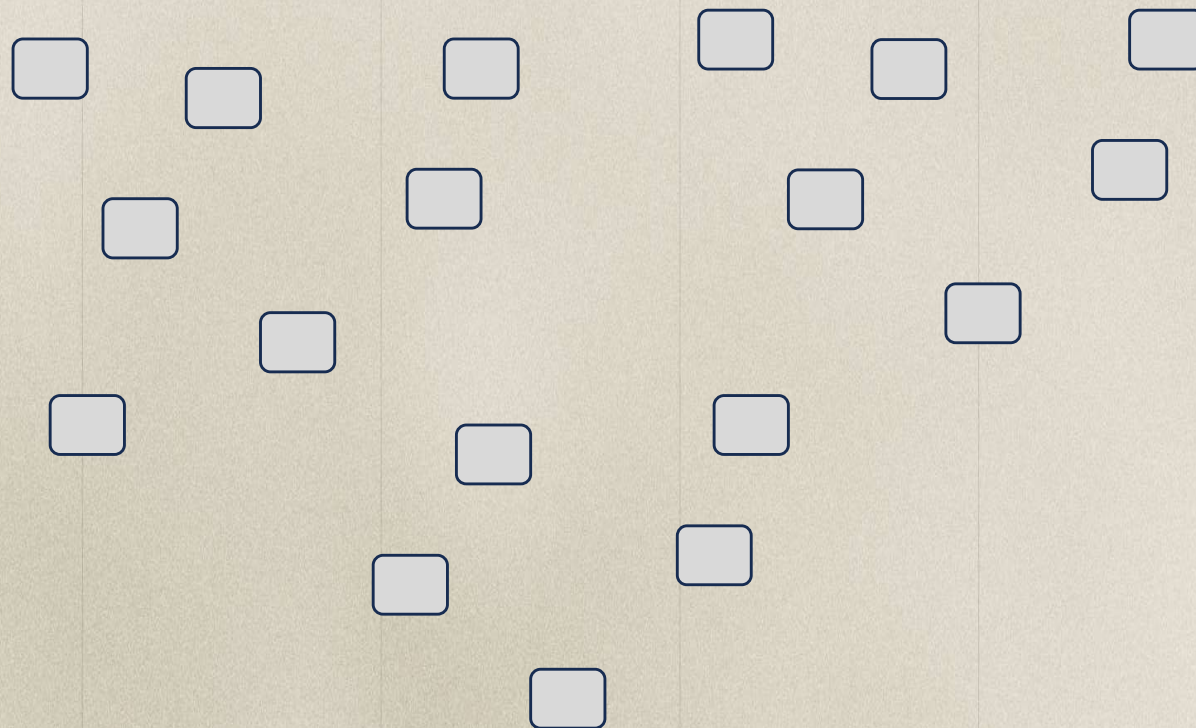


Como você resolveria?

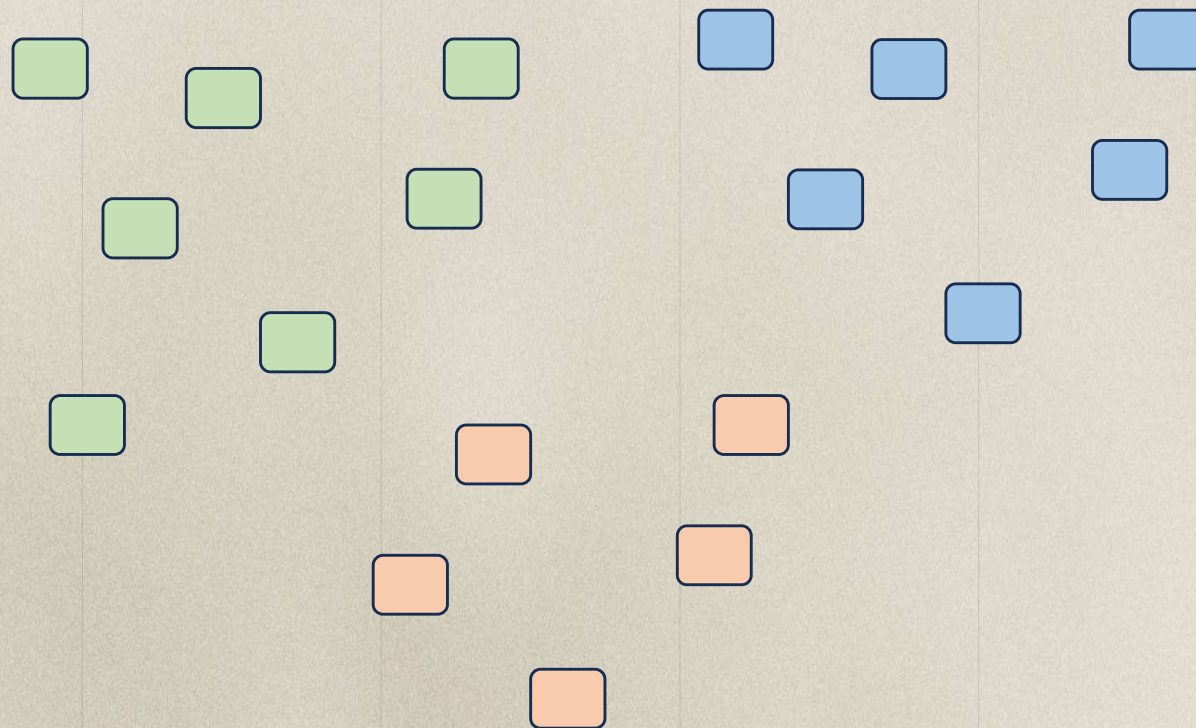
Uma empresa de manutenção de telefonia tem uma série de caixas centrais telefônicas numa grande cidade. Como dividir as n caixas em k áreas, para alocação de k técnicos?



K-Means

O K-Means é um algoritmo de clusterização

- Um cluster é um agrupamento
- O objetivo é particionar n observações dentre k grupos, onde cada observação pertence ao grupo mais próximo da média do cluster
- É um problema NP Hard



K-Means

O K-Means tem como objetivo:

- Minimizar variância dentro do cluster

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i$$

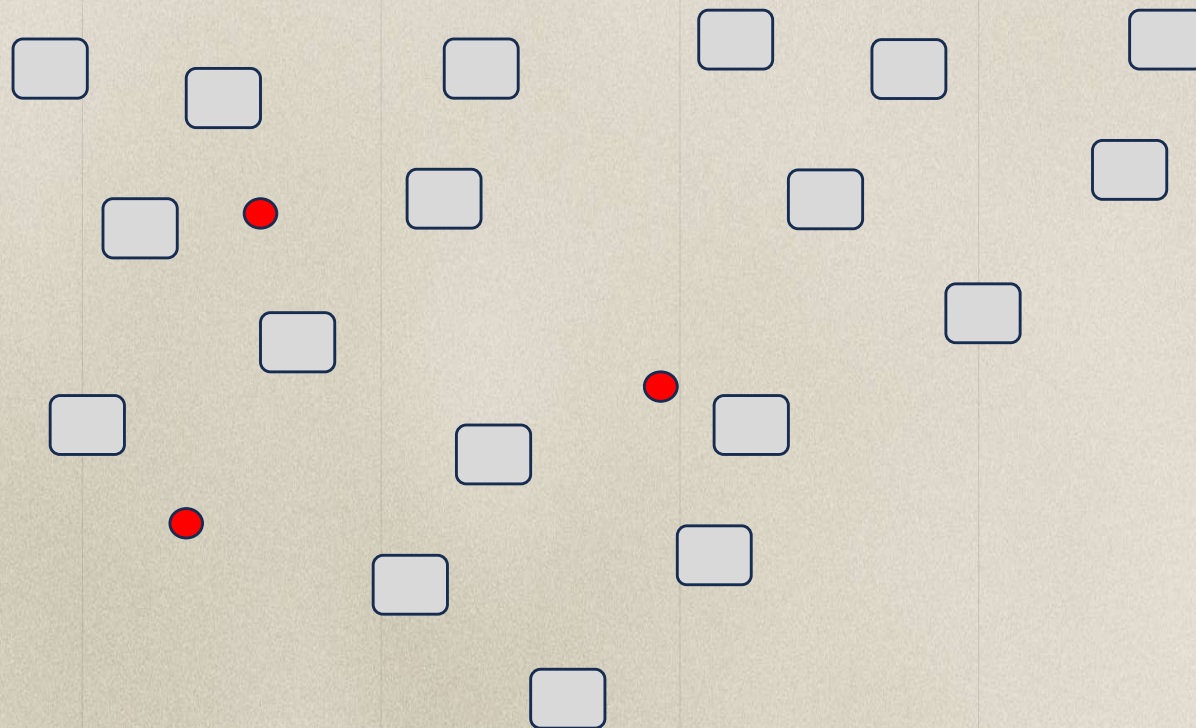
- A centroide é dada pela média dos pontos em cada cluster

$$\boldsymbol{\mu}_i = \frac{1}{|S_i|} \sum_{\mathbf{x} \in S_i} \mathbf{x},$$

K-Means

Algoritmo:

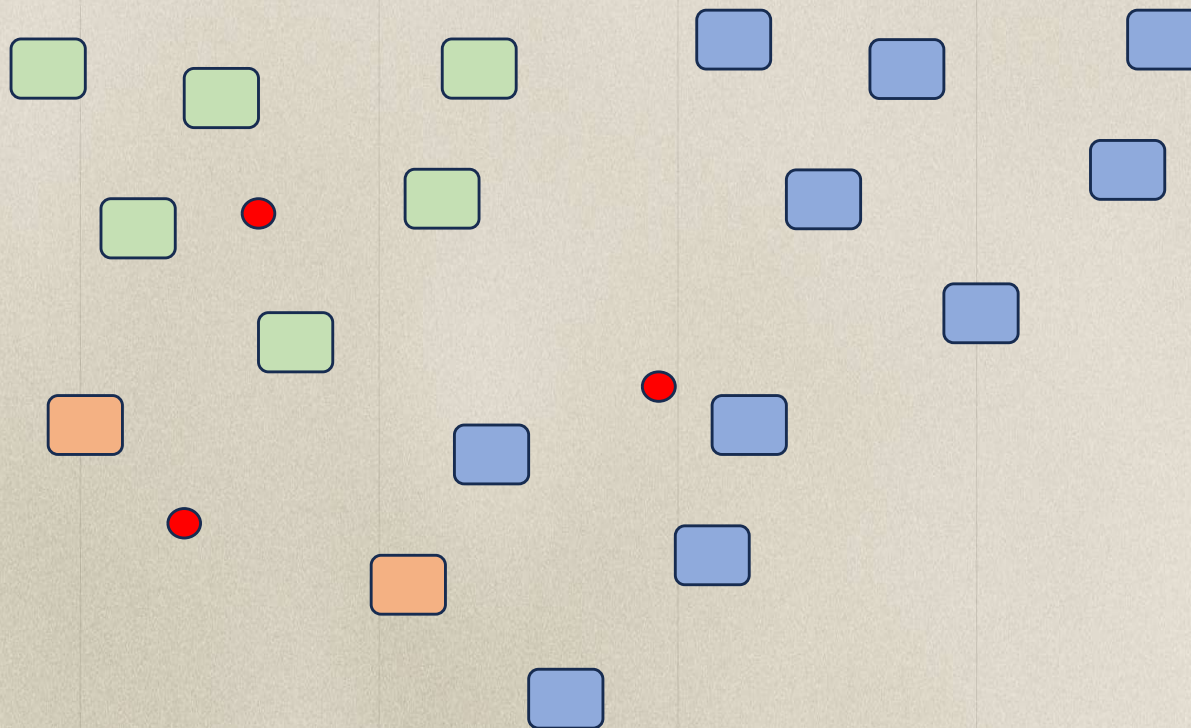
- Escolha k pontos aleatórios



K-Means

Algoritmo:

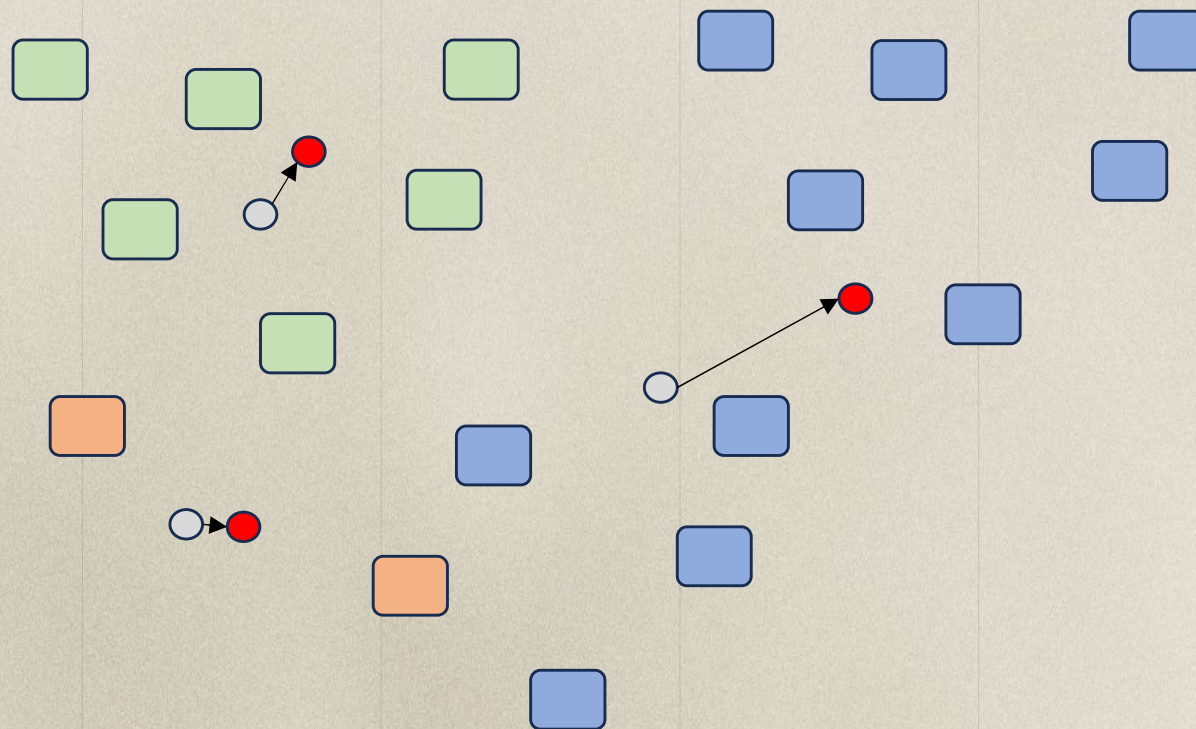
- Verifique quais os pontos mais próximos de cada centroide



K-Means

Algoritmo:

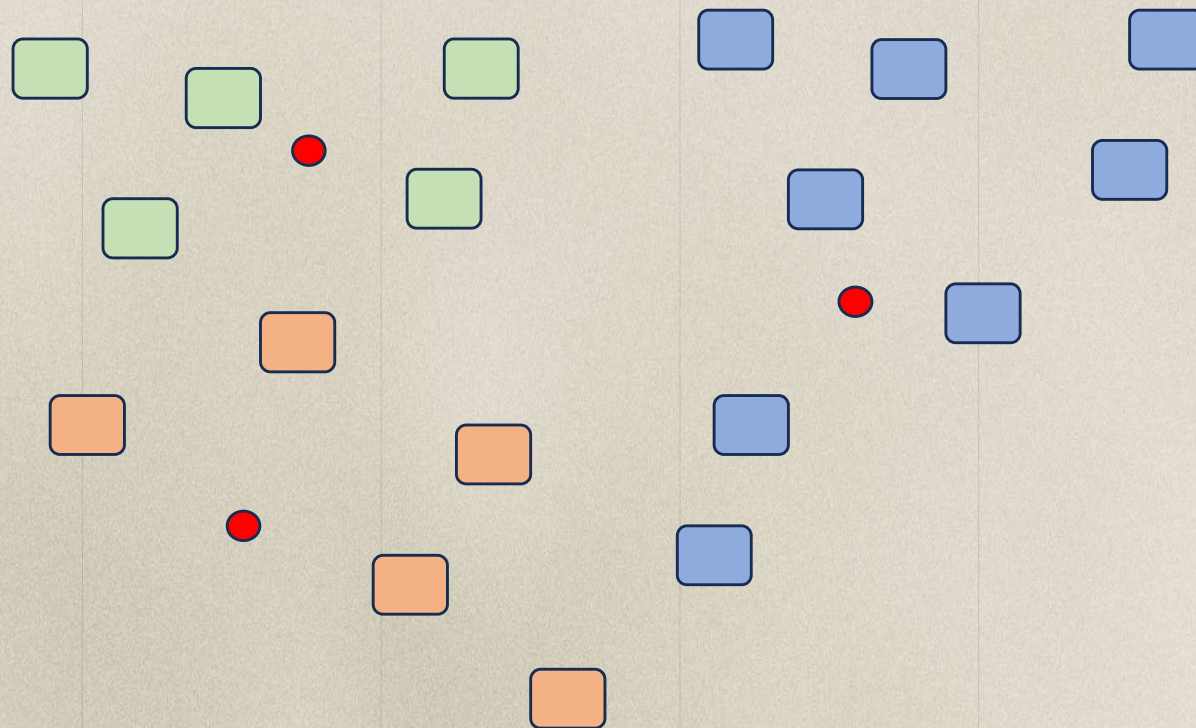
- Recalcule a centroide com os clusteres criados



K-Means

Algoritmo:

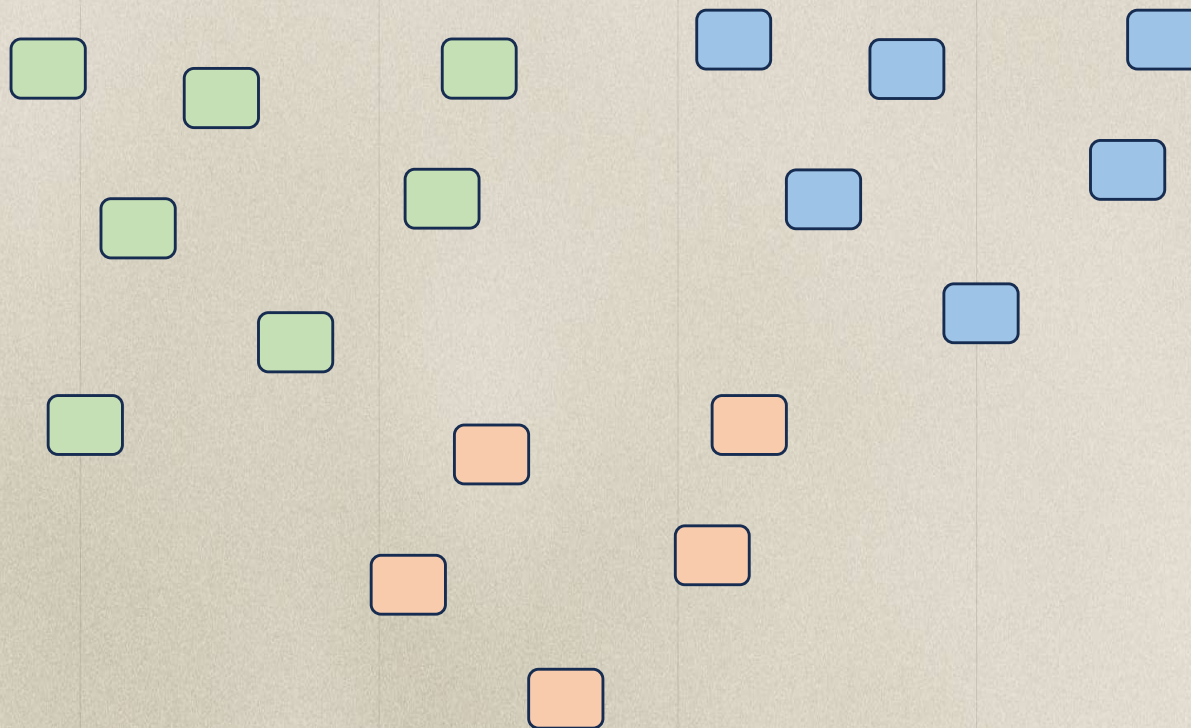
- Com a nova centroide, redistribua os pontos
- Repita até convergir para um resultado



K-Means

Algoritmo

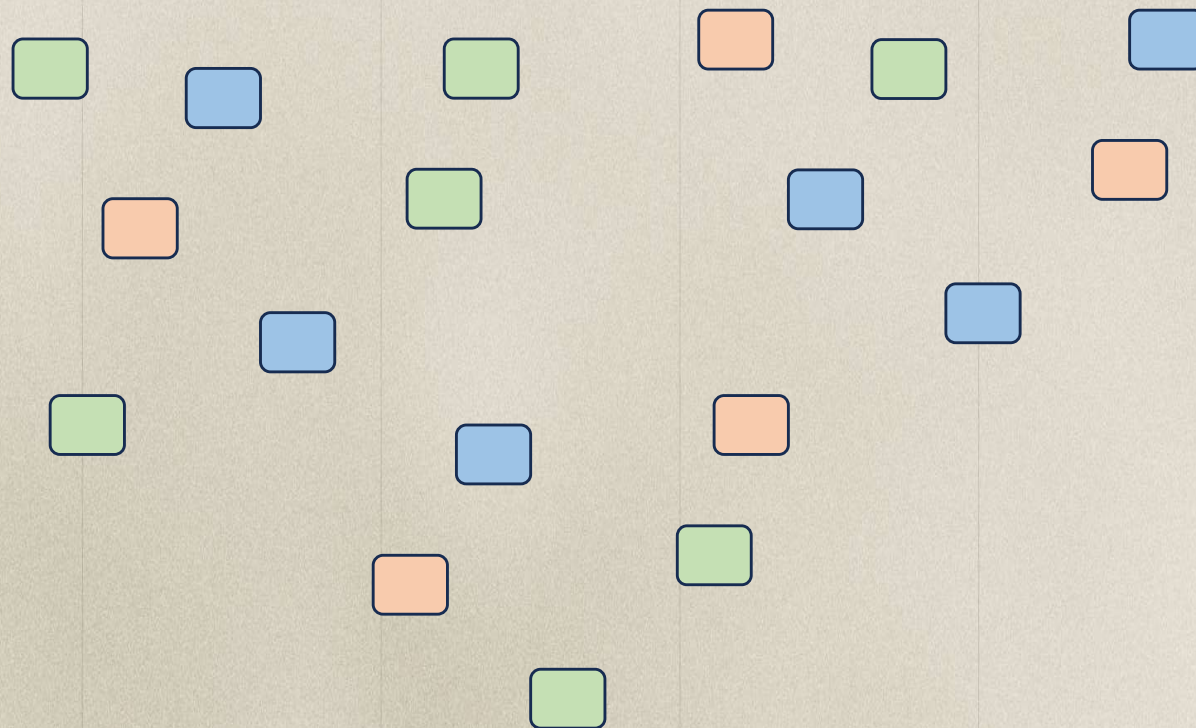
- Como a inicialização é aleatória, o resultado final pode diferir – portanto, é necessário repetir a operação até convergir num resultado



K-Means

Como fazer a inicialização?

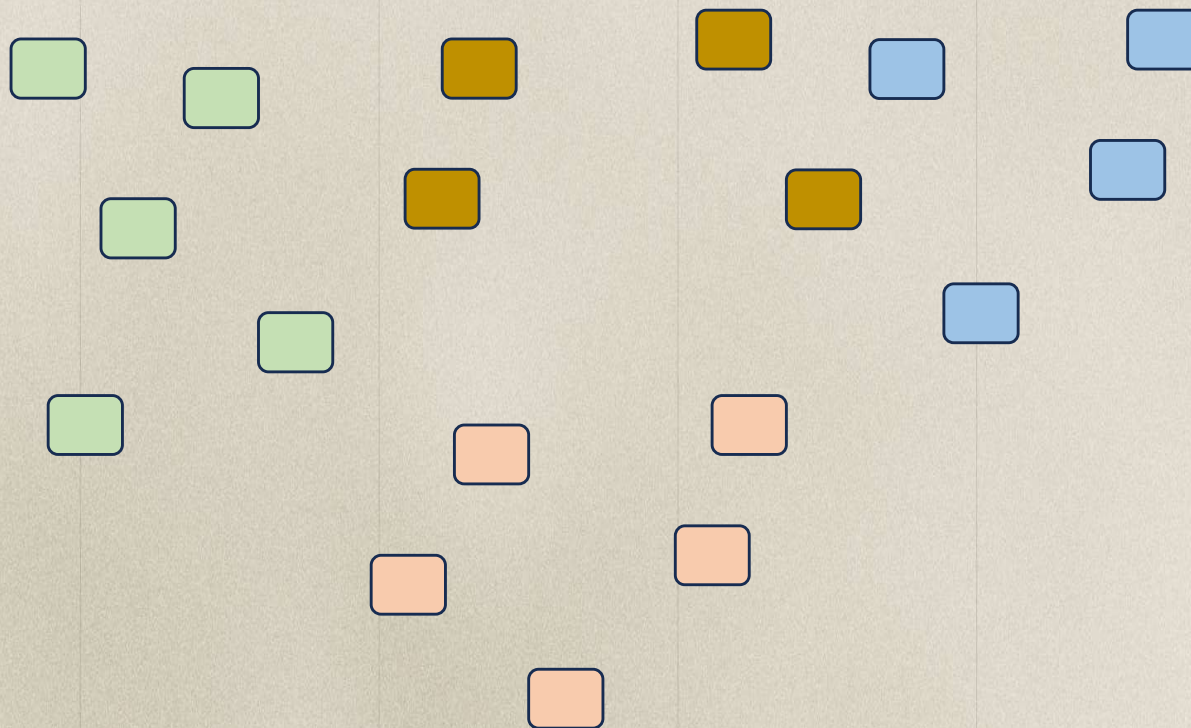
- Há outras formas de inicializar: ou pedir um chute inicial, ou escrever um centroide aleatório
- Uma heurística possível é o Random Partition: definir um cluster aleatório para cada ponto e inicializar o procedimento



K-Means

Quantos clusteres são necessários?

- É preciso definir de antemão o número de clusteres
- Pode vir de um input inicial, mas também pode vir de análise de sensibilidade, já que o algoritmo roda rapidamente: rodar com menos ou mais clusteres para analisar o resultado



K-Means

Conclusões

- O K-Means é um algoritmo bastante útil, e pode ser utilizado para problemas de classificação em geral
- É um exemplo de aprendizado não supervisionado, no mundo da IA

