

Aprendizado de Máquina e Deep Learning

K-means

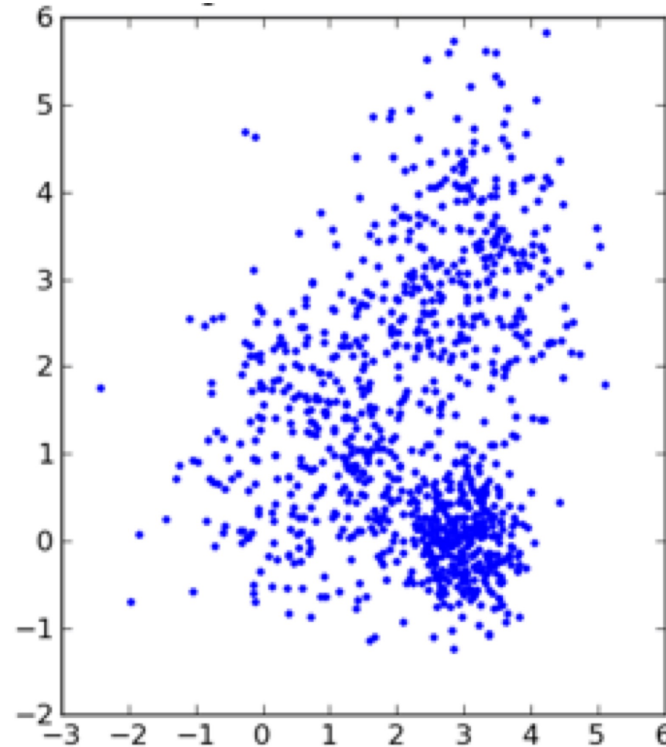
Prof. Dr. Thiago Meirelles Ventura

Agrupamento

- Uma das formas de aplicar aprendizado não supervisionado é dividir os dados em grupos
- Posteriormente, os dados contidos em determinado grupo representam uma informação

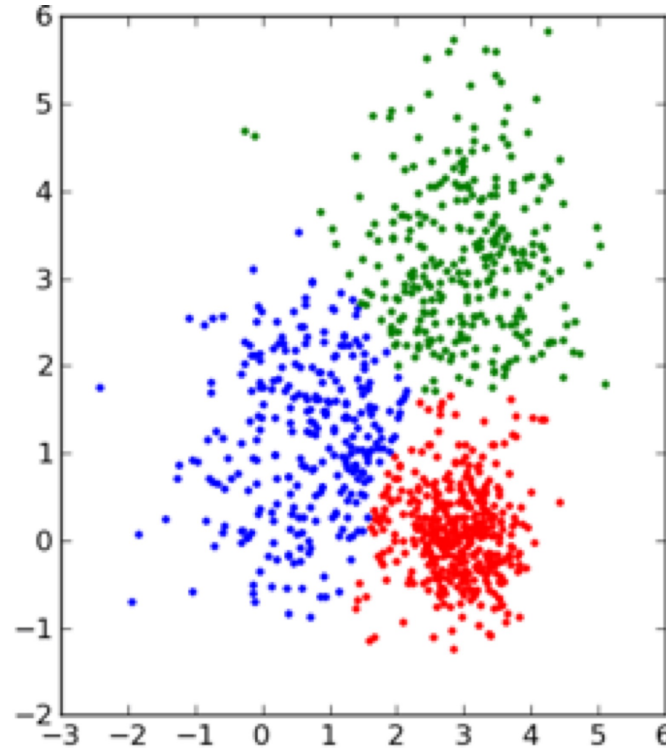
Agrupamento

- É possível dividir essa imagem em grupos?



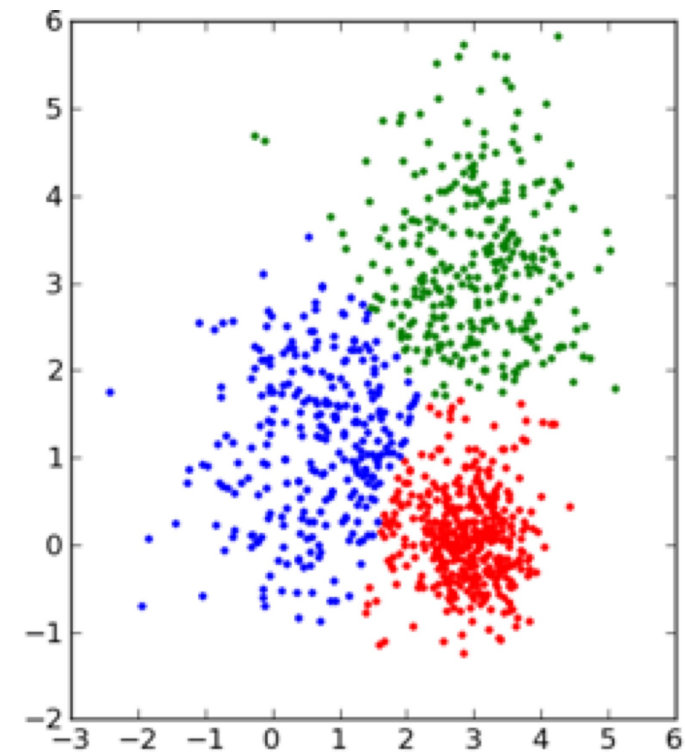
Agrupamento

- É possível dividir essa imagem em grupos?



Agrupamento

- Os dados foram separados em 3 grupos diferentes
- Isso significa que
 - Todos os elementos do grupo verde são similares entre si
 - O mesmo vale para os outros grupos
 - Ao surgir um novo exemplo, ele pode ser classificado em um dos 3 grupos e, assim, obter uma nova informação



Agrupamento

- Um sistema de recomendação pode se aproveitar desse conceito
- Imagine que, de acordo com as características de um filme, ele possa ser classificado em um dentre os vários grupos
- Caso você goste deste filme, é provável que você goste dos filmes que estão no mesmo grupo, uma vez que eles compartilham das mesmas características

Discussão

- Como podemos utilizar métodos de agrupamento para sugerir compras no Submarino?
- E para identificar qual espécie de ave pertence um determinado som?

Agrupamento

Mas como montar esses grupos?

K-means

- Um dos métodos não-supervisionados é o K-means
- Esse método divide os exemplos entre k grupos
- Pode ser calculado o centro de cada grupo (chamado de centróide)
- Cada exemplo da base de dados pertence ao grupo que possui o centróide mais próximo, de acordo com uma função de distância

K-means

- Primeiramente, é determinado um valor para k
 - quantos grupos serão montados
- Depois, são escolhidos k elementos que formarão os grupos iniciais
 - esses elementos representam o centróide do respectivo grupo

K-means

- Há várias formas de definir o centróide do grupo
- Uma maneira simples é calcular a média de cada característica de cada elemento pertencente ao grupo
- Como no início só haverá 1 elemento em cada grupo, ele acaba sendo o próprio centróide

K-means

- Em seguida, para cada exemplo é calculado a sua distância em relação ao centróide de cada grupo
- O exemplo é integrado ao grupo que possuir a menor distância
- Como o grupo foi alterado, deve ser recalculado o centróide deste grupo

K-means

- Depois que todos os exemplos foram inseridos em algum grupo
 - é verificado se algum exemplo precisa mudar de grupo
 - é recalcula a distância com os centróides atuais dos grupos
- O processo é finalizado quando
 - acontecem n iterações ou
 - quando não houver mais mudanças

K-means

- Resumo

1. Escolher k elementos que formarão os grupos iniciais
2. Calcular a distância de cada elemento em relação aos grupos
3. Agrupar o elemento ao grupo que possuir a menor distância
4. Recalcular o centróide do grupo
5. Repetir o processo para todos os outros exemplos
6. Verificar se cada elemento precisa mudar de grupo
7. Finalizar o processo se não houver mais mudança ou se n iterações forem realizadas

Exercício 3

- A classe será dividida em duas turmas: A e B
- Apenas a turma A responde o questionário:

<https://forms.gle/eRuq2xJZaxNrTff2A>



Exercício 4

- Sistema de indicações com estudantes
 - Cada estudante será um objeto da base de dados
 - Características: gosto por filmes, séries, jogos e livros
 - A sala será agrupada em 4 grupos

Exercício 4

- Passos

1. Listar 3 obras que goste (filmes, séries, jogos e/ou livros)
2. Montar o seu vetor de características
3. Aplicar K-means
4. Obter indicações

Exercício 4

- Listar 3 obras que goste (filmes, séries, jogos e/ou livros)
 - Escreva nas tiras de papéis o tipo da indicação e o nome
 - Exemplos:

“Filme - Guardiões da Galáxia: Volume 3”

“Série – The Office”

“Jogo – Zelda: tears of the Kingdom”

“Livro – O Código Da Vinci”

Exercício 4

- Montar o seu vetor de características
 - O vetor de características terá 7 valores
 - Avalie de 0 a 10 cada um dos gêneros a seguir de acordo com a sua preferência
 - Filmes ou séries de ação
 - Filmes ou séries de romance
 - Filmes ou séries de comédia
 - Jogos de puzzle
 - Jogos de esporte
 - Livros de fantasia
 - Livros de terror

Exercício 4

- Montar o seu vetor de características
 - O vetor de características terá 7 valores
 - Avalie de 0 a 10 cada um dos gêneros a seguir de acordo com a sua preferência
 - Filmes ou séries de ação 8
 - Filmes ou séries de romance 7
 - Filmes ou séries de comédia 9
 - Jogos de puzzle 4
 - Jogos de esporte 3
 - Livros de fantasia 5
 - Livros de terror 1

| | | | | | | |
|---|---|---|---|---|---|---|
| 8 | 7 | 9 | 4 | 3 | 5 | 1 |
|---|---|---|---|---|---|---|

Exercício 4

- Aplicar k-means
 - Decidir os 4 centróides
 - Cada estudante calcula a sua distância com cada grupo
 - Utilize a distância Manhattan
 - Vá para o grupo de menor distância
 - Recalculem o centróide do grupo (média entre todas as características)
 - Repitam
 - Calculem a sua distância com cada grupo
 - Mude de grupo
 - Recalculem o centróide do grupo

$$D(x, y) = \sum_{a=1}^n |x_a - y_a|$$

Exercício 4

- Obter indicações
 - Dividam o número de tiras pelo número de pessoas
 - Sorteiem as tiras entre si
 - Essas são indicações que talvez você goste

Pseudocódigo

```
kmeans (k, maxIteracoes)
  Selecione k elementos como centróides dos agrupamentos
  para todos os elementos restantes faça
    Calcule a distância entre ele e os centróides
    Adicione-o ao grupo que possuir a menor distância
    Recalcule o centróide do grupo
  fim para
  repita
    para todos os elementos faça
      Calcule a distância entre ele e os outros grupos
      Se houver uma distância menor que o grupo atual
        Mude-o de grupo
        Recalcule o centróide do novo grupo e do anterior
      fim se
    fim para
  até iterações > maxIteracoes ou não ocorra mudança
```

Exercício 5

- Apenas a turma B responde o questionário:

<https://forms.gle/eRuq2xJZaxNrTff2A>

