

Implementação do algoritmo CMeans

Inteligência Computacional

Sávio Rodrigues
Design de gestão e informática
CEFET-MG - Campus V
Divinópolis, Brasil
saviorrodrigues012@gmail.com

Resumo—O presente trabalho visa a implementação do algoritmo de agrupamento cmeans, supervisionado e não-supervisionado.

Palavras-chave—cmeans, agrupamento, supervisionado, não-supervisionado

I. INTRODUÇÃO

O cmeans é um algoritmo de agrupamento simples de ser implementado. Suas etapas de implementação são :

- Inicialização dos centros dos grupos.
- Determinar matriz de persistência.
- Calcular a função de custo.
- Atualizar os centros dos grupos.

II. DESENVOLVIMENTO

A base de dados foi criada de forma aleatória. Os pontos em cada grupo são definidos por um conjunto de valores randômicos que variam de 1 a 9. O critério de parada foi definido quando os centros não realocam mais ou quando se atinge 10 épocas.

A. Não supervisionado

No c-means não supervisionado, é calculada a distância dos pontos para todos os centros os quais são definidos aleatoriamente. Aquele que possui a menor distância é adicionado ao vetor de pertinência o valor do centro correspondente. Exemplo : Um ponto $x[i]$ pertence ao grupo de pontos próximos ao centro 2, portanto, o vetor de pertinência para a posição i é 2. Essa classificação é feita para todos os pontos obtendo como resultado um vetor de pertinência como mostrado na figura 1. É possível observar que se trata apenas de dois centros: 0 e 1.

[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.]

Fig. 1. Exemplo de um vetor de pertinência para 20 pontos

A função de custo é a média obtida de todos os pontos, para x e para y de um determinado grupo. O resultado é coletado para a definição do novo centro que será realocado para essa nova coordenada. Esse passo é repetido para todos os centros.

B. Supervisionado

No c-means supervisionado, os centros não são definidos de forma aleatória como no apresentado anteriormente. O ponto médio de cada grupo é valor inicial para aquele centro. O cálculo da função de custo e realocação de centros são os mesmos do algoritmo não-supervisionado.

Como é definido o centro de uma forma mais eficiente é esperado que leve menos épocas para que o algoritmo convirja.

III. RESULTADOS

Como mencionado anteriormente, foram definidos 2 grupos de 1000 pontos e dois centros. O agrupamento não supervisionado comportou-se como mostra a figura 3 em um total de 3 épocas. Importante observar que esse algoritmo nem sempre possui um resultado de saída bom. Isso acontece quando um dos centros sorteados fica distante dos dois grupos de dados, classificando-os apenas em um grupo.

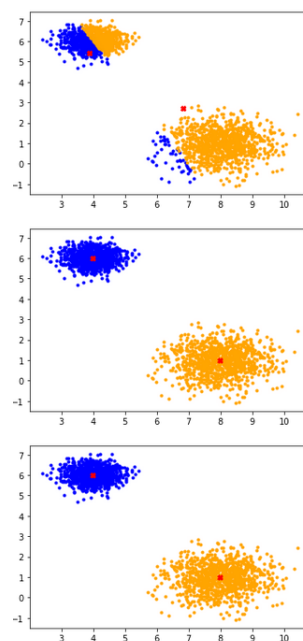


Fig. 2. Resultado c-means não supervisionado

Para a mesma base de dados o c-means supervisionado já convergiu em duas épocas - o mínimo para essa implementação.

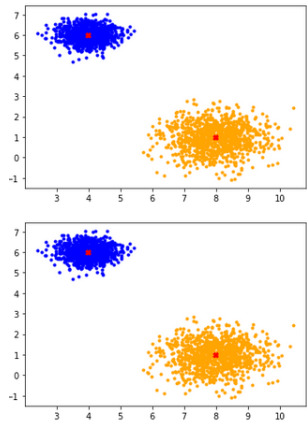


Fig. 3. Resultado c-means supervisionado

IV. CONCLUSÃO

O c-means supervisionado, em grande parte dos testes se mostrou mais eficiente ao convergir à um resultado satisfatório, menos custoso. Isso mostra o quanto é importante a inicialização dos centros para os resultados em custo, eficiência e cumprir com o objetivo de agrupar os dados.