



# ALGORITMOS GENÉTICOS PARA AGRUPAMENTO DE DADOS DE COBERTURA VEGETAL



Orientador: Prof. Dr. Francisco A. M. Gomes  
[chico@ime.unicamp.br](mailto:chico@ime.unicamp.br)

Alexandre Lorca Serafini  
[Alexandre\\_lorca@hotmail.com](mailto:Alexandre_lorca@hotmail.com)

Instituto de Matemática, Estatística e Computação Científica – IMECC  
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Palavras-chave: Algoritmo genético - Médias-k - Agrupamento de dados

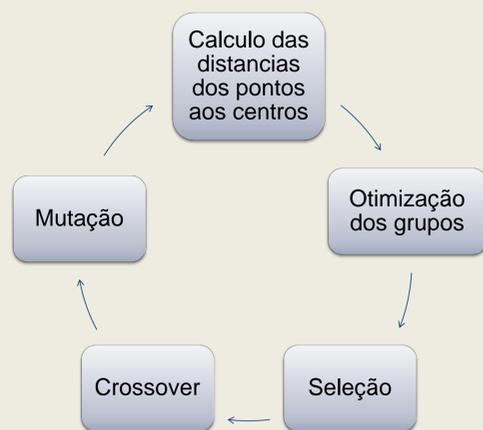
## Resumo

Este projeto visa identificar áreas do estado de São Paulo que tenham um mesmo tipo de vegetação. Essa identificação é feita determinando-se, a partir de imagens de satélite, as áreas que possuem “níveis” de NDVI similares através de um algoritmo genético de agrupamento de dados. Assim, áreas que tenham plantio, por exemplo, de feijão apresentarão níveis próximos de NDVI e farão parte de um mesmo grupo.

Para garantir que não haja nenhuma interferência no processo de agrupamento dos dados, todos os processos que o algoritmo realiza são gerados aleatoriamente. O programa utilizado para realizar esta tarefa foi o MATLAB, e que demonstrou ótimos resultados no processo de desenvolvimento do algoritmo

## Algoritmo Genético

Um algoritmo é uma série de comandos para executar um objetivo. Em especial, o algoritmo genético é assim chamado porque os passos por ele executados se assemelham à teoria da seleção natural. O esquema abaixo exemplificará melhor as atividades realizadas por este tipo de algoritmo para o problema de agrupamento de dados.



Os dados com os quais trabalhamos são os  $k$  níveis de NDVI de uma série histórica. Assim, cada um dos  $N$  grupos nos quais dividimos os dados é um conjunto de pontos no espaço de dimensão  $k$ . Um grupo é representado por seu centro geométrico, de modo que cada “cromossomo” é o conjunto de  $N$  centros. Inicialmente, escolhemos aleatoriamente  $N$  pontos para compor um cromossomo.

Na otimização dos grupos, calculamos as distâncias de todos os pontos aos centros, e reagrupamos os pontos de modo que cada centro esteja a uma distância mínima dos pontos que pertencem ao seu grupo.

Para a seleção, é preciso calcular a aptidão de cada cromossomo, que é dada pela soma das distâncias dos pontos aos centros dos grupos desse cromossomo. Assim, o melhor cromossomo é aquele que tem função de aptidão mais baixa. A seleção é feita pelo método da roleta, no qual cada cromossomo tem uma chance de permanecer na população proporcional ao inverso de sua aptidão. Assim, são eliminados cromossomos menos aptos, e os melhores podem ser até duplicados.

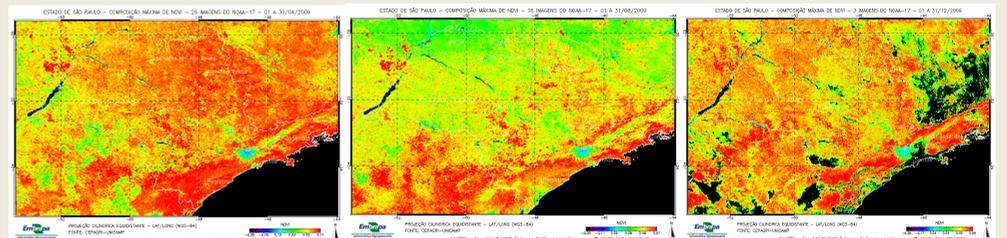
Para evitar que o algoritmo caia em um mínimo local, ou seja, que a resposta não seja a melhor possível, já que o algoritmo não consiga avançar com os dados que possui, são tomadas duas precauções. Uma delas é o *crossover*, que combina informações de um cromossomo com o de outro cromossomo. A outra é a *mutação*, que tem a tarefa de realizar uma pequena alteração nos centros. Variando a solução dessa maneira, o algoritmo é capaz de buscar uma solução em regiões distantes, porém promissoras.

## NDVI

NDVI (Normalized Difference Vegetation Index) ou IVDN (Índice de Vegetação por Diferença Normalizada) foi o meio utilizado para identificar os diferentes tipos de cobertura vegetal, este índice mede a diferença de refletância entre o infravermelho e o vermelho, para tal é realizada a seguinte equação:

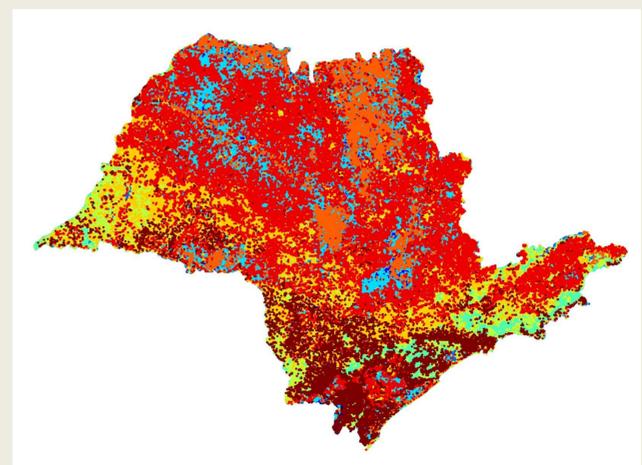
$$NDVI = \frac{IVP - V}{IVP + V}$$

Abaixo estão três imagens correspondentes respectivamente aos meses de abril, agosto e dezembro do ano de 2009, estas fotos foram captadas pelo satélite da Embrapa NOAA-17, as fotos retrataram melhor como o índice é captado.



## Resultados numéricos

O algoritmo foi usado para dividir em 10 grupos a cobertura do solo do estado de São Paulo. Para tanto, foram usados os valores do NDVI coletados pelo satélite do NOAA nos 12 meses de 2009. Após 100 iterações do algoritmo genético, aplicou-se o método das médias-k para otimizar o resultado. Através das ferramentas gráficas do MATLAB, foi possível desenhar o mapa abaixo, com o resultado final.



## Conclusões

O algoritmo genético usado conseguiu realizar o agrupamento dos dados proposto, de modo que essa estratégia pode ser aplicada para identificar com certa precisão o tipo de cobertura vegetal de uma região a partir dos dados de NDVI captados por satélite.

## Referências

- [1] BANDYOPADHYAY, S. & MAULIK, U. – An evolutionary technique based on K-means algorithm for optimal clustering in  $R^n$ . Information Sciences 146, p. 221–37, 2002.
- [2] CHANG, D. X.; ZHANG, X. D.; ZHENG, C. W. – A genetic algorithm with gene rearrangement for K-means clustering. Pattern Recognition 42, p. 1210–22, 2009.