

# Segmentação de Dados Sísmicos Utilizando K-means

---

Aplicação de Aprendizado Não Supervisionado em Processamento  
Sísmico

Autores: **José Augusto** e **Ernane Ferreira**

Análise de Sismogramas e Processamento de Dados Sísmicos

2025

# O Algoritmo K-means: Conceitos Fundamentais

Entendendo o processo de convergência através das iterações

## O que é K-means?

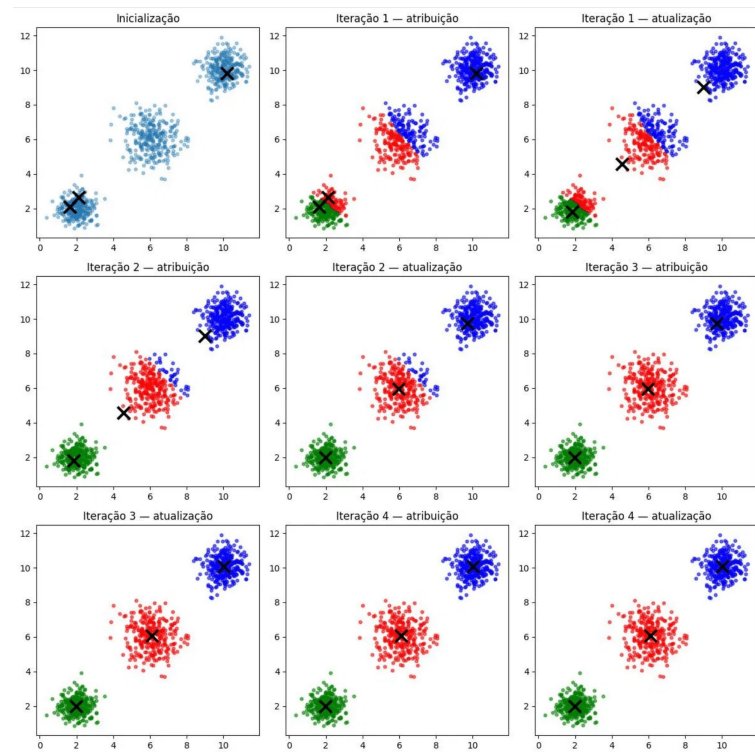
K-means é um algoritmo de aprendizado não supervisionado que particiona dados em **k clusters** distintos, minimizando a variância dentro de cada grupo.

## Etapas do Algoritmo

- ▶ **Inicialização:** Seleção aleatória de k centróides
- ▶ **Atribuição:** Cada ponto é associado ao centróide mais próximo
- ▶ **Atualização:** Centróides são recalculados como média dos pontos do cluster
- ▶ **Convergência:** Processo se repete até estabilização

## Aplicação em Sísmica

Utiliza-se K-means para segmentar dados sísmicos em padrões distintos, permitindo identificar sinal, ruído e estruturas geológicas.



# Contexto da Pesquisa em Sísmica

Importância da segmentação em dados sísmicos

## Desafios em Processamento Sísmico

Dados sísmicos são complexos e contêm múltiplas escalas de informação. A presença de ruído, artefatos e variações de amplitude dificultam a interpretação automática.

## Por que Segmentar?

- ▶ Distinguir **signal de ruído**
- ▶ Identificar **padrões geológicos**
- ▶ Melhorar **interpretação** de estruturas
- ▶ Automatizar **processamento** em larga escala

## Abordagem Não Supervisionada

K-means permite segmentação **sem rótulos prévios**, descobrindo padrões naturais nos dados sísmicos.

## Aplicações Práticas

- Exploração de petróleo
- Mapeamento geológico
- Avaliação de reservatórios
- Deteção de anomalias sísmicas

## Objetivo desta Pesquisa

Aplicar K-means com múltiplos atributos sísmicos para segmentação automática de sismogramas, demonstrando viabilidade de aprendizado não supervisionado em processamento sísmico.

# Metodologia: Pipeline de Processamento

Etapas sequenciais de processamento sísmico e segmentação

1

## Carregamento de Arquivo SEG-Y

Leitura do arquivo sísmico em formato SEG-Y (Jequitinhonha). Extração de traces e amostras de tempo.

*Dados: 12.060 traces  $\times$  1.001 amostras (linha 2D)*

2

## Extração de Atributos Sísmicos

Cálculo de **3 atributos principais**: envelope, amplitude e fase.

*Matriz de features: 12.072.060 voxels  $\times$  3 atributos*

3

## Normalização de Dados

Padronização dos dados utilizando StandardScaler (média 0, desvio padrão 1).

*Essencial para garantir que todos os atributos tenham igual peso no K-means*

4

## Aplicação do K-means

Segmentação com **k=6 clusters** (determinado por Silhouette Score). Inicialização k-means++ com 10 iterações.

*Amostragem de 1% dos dados para avaliação do número ótimo de clusters*

# Atributos Sísmicos Selecionados

Três dimensões de informação para segmentação robusta

## Envelope

Magnitude do Sinal Analítico

### Cálculo

$$\text{Env} = \sqrt{s^2 + H(s)^2}$$

Onde  $s$  é o sinal original e  $H(s)$  é a transformada de Hilbert

### Importância

- ✓ Captura **energia** do sinal
- ✓ Independente de **fase**
- ✓ Robusto a **variações** de polaridade

## Amplitude

Magnitude do Sinal

### Cálculo

$$\text{Amp} = |s(t)|$$

Valor absoluto do sinal sísmico em cada amostra

### Importância

- ✓ Reflete **contraste** de impedância
- ✓ Indicador de **força** do sinal
- ✓ Sensível a **mudanças** litológicas

## Fase

Ângulo do Sinal Analítico

### Cálculo

$$\phi = \arctan(H(s)/s)$$

Ângulo entre componentes real e imaginária

### Importância

- ✓ Detecta **continuidade** lateral
- ✓ Identifica **descontinuidades**
- ✓ Independente de **amplitude**

# Seleção do Número Ótimo de Clusters (k=6)

Método de Silhouette Score para determinar k ideal

## Método Silhouette Score

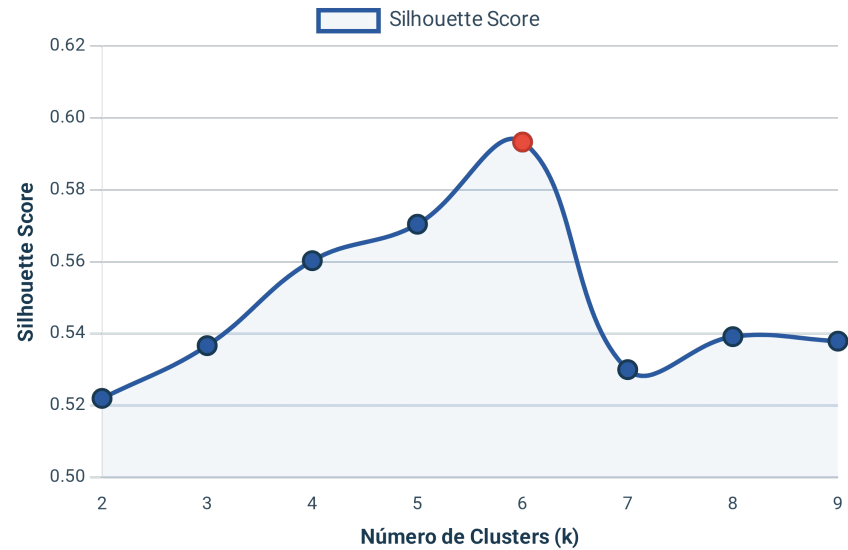
O Silhouette Score mede a qualidade da segmentação, calculando o quão bem cada ponto se ajusta ao seu cluster em relação aos demais. Valores variam de -1 a 1, onde valores maiores indicam melhor segmentação.

## Estratégia de Avaliação

Para acelerar a avaliação, utilizou-se uma **amostra de 1%** dos dados (120.720 voxels de 12.072.060 totais). Testaram-se valores de k de 2 a 9.

## Resultados

K	Score
2	0.5228
3	0.5380
4	0.5625
5	0.5730
6	0.5967
7	0.5311
8	0.5406
9	0.5393



# Resultados: Segmentação com k=6

Distinção entre sinal sísmico e ruído

## Parâmetros da Segmentação

- Número de Clusters: k = 6
- Silhouette Score: 0.5967
- Dados Utilizados: 1% da amostra (120.720 voxels)
- Total Classificado: 12.072.060 voxels

## Qualidade da Segmentação

O Silhouette Score de **0.5967** indica uma segmentação de boa qualidade, com clusters bem definidos e separação clara entre grupos.

## Interpretação dos Clusters

- Clusters 0 e 1: Ruído
- Cluster 2: Sinal intermediário
- Clusters 3, 4 e 5: Ruído estruturado

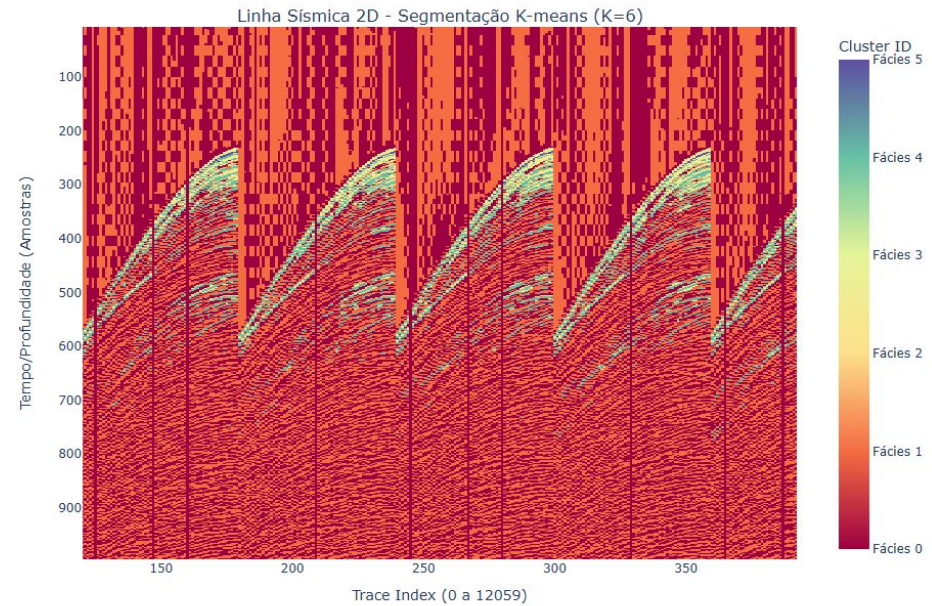
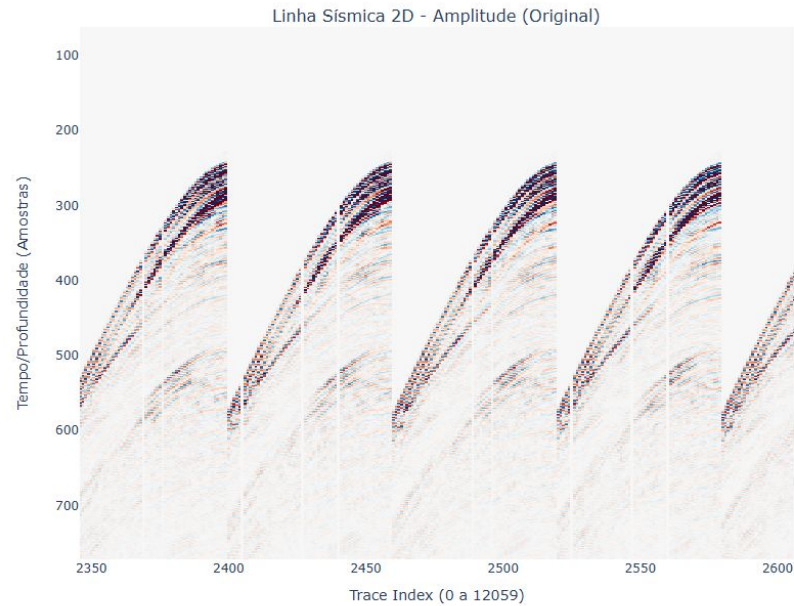
## Principais Resultados

- Segmentação automática bem-sucedida
- Distinção entre **sinal e ruído**
- Clusters estáveis e interpretáveis
- Aplicável a todo volume sísmico

# Resultados: Segmentação com k=6

Distinção entre sinal sísmico e ruído

Segmentação de Fácies Sísmicas (Linha 2D Completa)





# Próximos Passos: Otimização de Features

Redução de dimensionalidade e combinações de atributos

## Features Atuais (3 atributos)

Atributo	Descrição
Envelope	Valor real + Hilbert
Amplitude	Magnitude do sinal
Fase	Informação de fase

## Objetivo da Otimização

Identificar qual combinação de features oferece melhor desempenho com **menor complexidade computacional**.

## Combinações a Testar (2 features)

<b>1. Envelope + Amplitude</b> Combina energia e magnitude do sinal
<b>2. Envelope + Fase</b> Combina energia com informação de fase
<b>3. Amplitude + Fase</b> Combina magnitude com fase do sinal

## Métricas de Avaliação

- Silhouette Score para cada combinação
- Tempo de processamento
- Qualidade da segmentação
- Comparação com k=6 (baseline)

# Próximos Passos: Dados Pós-stack

Validação da metodologia em dados empilhados

## Pré-stack vs Pós-stack

Característica	Pré-stack	Pós-stack
Ruído	Maior	Menor (empilhado)
Informação	Completa	Condensada
Processamento	Intensivo	Eficiente

## Justificativa

Testar em dados pós-stack permite validar a robustez da metodologia K-means em dados com melhor razão sinal/ruído e menor dimensionalidade.

## Estratégia de Teste

<b>1. Carregamento de Dados Pós-stack</b>
Utilizar arquivo SEG-Y pós-stack correspondente
<b>2. Extração de Atributos</b>
Aplicar os teste de atributos (envelope, amplitude, fase) citados aos dados pós-stack
<b>3. Avaliação de k</b>
Repetir análise de Silhouette Score para determinar k ótimo em dados pós-stack
<b>4. Comparação de Resultados</b>
Contrastar segmentações pré-stack vs pós-stack

## Métricas Esperadas

- Silhouette Score pós-stack
- Número ótimo de clusters
- Qualidade da segmentação

# Próximos Passos: Análise por CMP

Investigação de padrões localizados em escala reduzida

## O que é CMP?

**CMP (Common Midpoint)** é um agrupamento de traços sísmicos que compartilham o mesmo ponto médio entre fonte e receptor. Fundamental em processamento sísmico.

## Por que Analisar CMPs?

- Reduz escala de dados para análise detalhada
- Permite investigação de **padrões localizados**
- Facilita validação de hipóteses
- Melhor compreensão do comportamento do K-means

## Escala de Dados

Tipo	Traces	Amostras
<b>Linha Completa</b>	12.060	1.001
<b>CMP Único</b>	~50-100	1.001

## Objetivos da Análise

### 1. Validação Local

Confirmar que a segmentação funciona bem em dados reduzidos

### 2. Padrões Estruturais

Investigar como os clusters se distribuem dentro de um CMP

### 3. Variabilidade Lateral

Analisar mudanças de amplitude e fase ao longo do CMP

### 4. Qualidade de Segmentação

Avaliar Silhouette Score em dados localizados

## Metodologia

- Selecionar **1 CMP representativo**
- Extrair atributos (envelope, amplitude, fase)
- Aplicar K-means variando os atributos e descobrir o melhor K
- Visualizar segmentação local
- Comparar com resultados globais

# Próximos Passos: Visualizações Avançadas

Análise detalhada de features, clusters e relações 3D

## 1. Plotagem Individual de Features

### Envelope (Real + Hilbert)

Visualização em 2D da distribuição espacial do envelope sísmico

*Mostra energia do sinal em toda a linha sísmica*

### Amplitude

Mapa de amplitude mostrando variações de magnitude

*Identifica zonas de forte contraste de impedância*

### Fase

Distribuição espacial da informação de fase

*Revela continuidade lateral e descontinuidades*

## 2. Plotagem Individual de Clusters

Cada um dos **6 clusters** será visualizado isoladamente em um mapa 2D, permitindo análise detalhada de sua distribuição espacial.

Cluster 0	Cluster 1
Cluster 2	Cluster 3
Cluster 4	Cluster 5

Cada mapa mostrará apenas os voxels pertencentes ao cluster específico, facilitando a interpretação de padrões localizados.

## 3. Visualização 3D dos Clusters

Plotagem no espaço tridimensional dos **3 atributos** (Envelope, Amplitude, Fase), permitindo visualizar a relação entre os clusters em 3D.

Cada ponto representa um voxel

Cores indicam o cluster

Visualização interativa

Análise de separabilidade

# Conclusões e Perspectivas

Síntese dos resultados e potencial da abordagem

## Principais Resultados Obtidos

### Segmentação Bem-sucedida

Aplicação bem-sucedida de K-means em dados sísmicos com **k=6 clusters** e Silhouette Score de **0.5967**.

### Distinção Sinal-Ruído

Segmentação automática permitiu **distinguir claramente** sinal sísmico de ruído em toda a linha 2D.

### Escalabilidade

Metodologia aplicada com sucesso a **12.072.060 voxels** do volume sísmico completo.

## Contribuições da Pesquisa

Metodologia robusta com 3 atributos sísmicos

Abordagem não supervisionada viável

Pipeline reproduzível e escalável

## Aplicações Práticas em Exploração

### Mapeamento Geológico

Identificação automática de estruturas e padrões geológicos

### Avaliação de Reservatórios

Segmentação de zonas produtivas e não produtivas

### Deteção de Anomalias

Identificação de padrões anômalos e estruturas especiais

### Processamento em Larga Escala

Automatização de análise de múltiplas linhas sísmicas

## Perspectivas Futuras

Otimização de features (2 atributos)

Validação em dados pós-stack

Análise detalhada por CMP

Visualizações 3D avançadas

Integração com workflows de interpretação