

Clusterização de Dados Utilizando MATLAB®

Kenneth Brenner dos Anjos Benício - 519189

Universidade Federal do Ceará - Mestrado em Engenharia de Teleinformática

Resumo

Neste trabalho, foram desenvolvidas simulações computacionais utilizando MATLAB® para que fossem estudados os desempenhos de dois algoritmos de clusterização considerando um conjunto de dados qualquer, visto que não existem informações disponíveis sobre tal conjunto. Por fim, foi obtido um valor ótimo de clusterização para o conjunto de dados, além de ter sido demonstrado matematicamente que existe uma equivalência entre os dois algoritmos de clusterização utilizados no trabalho.

1 Introdução

O relatório é dividido em três seções: Na primeira seção abordo as formulações e procedimentos matemáticos que foram utilizadas na geração dos resultados. Em seguida, apresento o procedimento para a escolha do valor ótimo para o número de agrupamentos, K , além de também exibir algumas estatísticas mais fundamentais a respeito dos protótipos e agrupamentos após a definição de um valor para K . Por fim, na última seção discuto brevemente minhas conclusões sobre os resultados obtidos.

1.1 Conjunto de Dados

O conjunto de dados foi simplesmente fornecido pelo professor da disciplina sem maiores detalhes sobre a significância dos dados processados.

1.2 Convergência do Algoritmo

Antes de seguir com as definições dos algoritmos que serão utilizados é interessante definir o conceito de Soma das Distâncias Quadráticas (SSD). O SSD é uma expressão utilizada para avaliar o posicionamento dos protótipos durante a execução dos algoritmos de clusterização e é definido como

$$SSD(K) = \sum_{i=1}^K \sum_{\forall \mathbf{x} \in V_i} \|\mathbf{x} - \mathbf{w}_i\|^2. \quad (1)$$

Desse modo, essa será a métrica utilizada ao longo deste trabalho para avaliar o processo de convergência dos algoritmos de clusterização.

1.3 K-Means Sequencial

O K-Means sequencial é definido no Algoritmo 1

Algorithm 1 Algoritmo K-Means Sequencial

Passo 1: Definir um valor para o hiperparâmetro K .
Passo 2: Inicializar os K protótipos.
Passo 3: Determinar o índice do protótipo mais próximo ao vetor de atributos atual pela expressão

$$i'(t) = \arg \min_{\forall i} \|\mathbf{x}(n) - \mathbf{w}_i(n)\|^2 \quad (2)$$

Passo 4: Atualizar a posição do protótipo $\mathbf{w}_{i'}$ de acordo com a expressão recursiva abaixo

$$\mathbf{w}_{i'}(n+1) = \mathbf{w}_{i'}(n) + \alpha_{i'}(n) [\mathbf{x}(n) - \mathbf{w}_{i'}(n)]. \quad (3)$$

Passo 5: Repetir os passos 3-4 enquanto não ocorrer a convergência das posições dos protótipos.

1.4 K-Means em Batch

O K-Means em Batch é definido no Algoritmo 2

Algorithm 2 Algoritmo K-Means em Batch

Passo 1: Definir um valor para o hiperparâmetro K .
Passo 2: Inicializar os K protótipos.
Passo 3: Determinar a partição V_i a qual pertence o parâmetro \mathbf{w}_i , onde $i = 1, \dots, K$, com o uso da expressão

$$V_i = \{\mathbf{x} \in \mathbb{R}^p \mid \|\mathbf{x} - \mathbf{w}_i\|^2 < \|\mathbf{x} - \mathbf{w}_j\|^2, \forall j \neq i\}. \quad (4)$$

Passo 4: Atualizar a posição do protótipo \mathbf{w}_i de acordo com a média de objetos presentes na partição V_i

$$\mathbf{w}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in V_i} \mathbf{x}. \quad (5)$$

Passo 5: Repetir os passos 3-4 enquanto não ocorrer a convergência das posições dos protótipos.

1.5 A equivalência entre os métodos

É relevante demonstrar também que os protótipos convergem para os vetores centróides do conjunto de dados a medida que as iterações avançam. Se considerarmos que a expressão definida em (2) é um processo estacionário no sentido amplo (WSS) então, após aplicarmos o operador média em ambos os lados, podemos escrever $\mathbb{E}\{\mathbf{w}_{i'}(n+1)\} - \mathbb{E}\{\mathbf{w}_{i'}(n)\} = 0$ e chegar na seguinte reescrita para a expressão de atualização do K-Means sequencial

$$\begin{aligned}\alpha_{i'} \mathbb{E}\{\mathbf{x}(n) - \mathbf{w}_{i'}(n)\} &= 0, \\ \alpha_{i'} \mathbb{E}\{\mathbf{w}_{i'}(n)\} &= \alpha_{i'} \mathbb{E}\{\mathbf{x}(n)\}, \\ \mathbb{E}\{\mathbf{w}_{i'}(n)\} &= \frac{1}{N_i} \sum_{\mathbf{x} \in V_i} \mathbf{x}(n).\end{aligned}$$

Portanto, é possível transitar da equação recursiva da atualização dos protótipos para o K-Means sequencial até a equação que define a atualização dos protótipos para o K-Means em Batch.

1.6 Software

Todos os códigos foram desenvolvidos utilizando-se o MATLAB[®] 2021a. Foi criada uma classe de funções com métodos correspondentes aos classificadores aqui abordados. Todos os códigos devidamente comentados foram enviados conjuntamente com esse relatório. Ademais, os resultados foram gerados por um computador com processador Intel i7-10700K (3.8GHz), 16GB de memória RAM e uma placa gráfica RX 6600.

2 Resultados

- Antes de prosseguir é necessário definir as expressões matemáticas para os índices que serão utilizados para a análise numérica do desempenho. As métricas utilizadas tentam avaliar duas principais características do particionamento: A coesão interna, objetos de um mesmo agrupamento deveriam ser tão similares entre si quanto possível, e a separação externa, agrupamentos deveriam estar tão distantes entre si quanto possível. Desse modo, é possível definir as três métricas abaixo que irão avaliar tais aspectos

$$\begin{aligned}\text{Dunn} &= \frac{\min_{i \neq j} \{\delta(V_i, V_j)\}}{\max_{1 \leq l \leq K} \{\Delta(V_l)\}}, \\ \text{Davies-Bouldin} &= \frac{1}{K} \sum_{i=1}^K \max_{j, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right\}, \\ \text{Calinski-Harabasz} &= \frac{N - K}{K - 1} \frac{\text{tr}(\mathbf{S}_B)}{\text{tr}(\mathbf{S}_w)}.\end{aligned}$$

Para o índice de Dunn $\delta(V_i, V_j)$ indica uma métrica de dissimilaridade entre diferentes partições, $\Delta(V_l)$ indica uma métrica de dispersão interna de dados para uma partição l . Já para o índice de Davies-Bouldin $S_{i,q} = \left[\frac{1}{N_i} \sum_{\mathbf{x} \in V_i} \|\mathbf{x} - \mathbf{w}_i\|^{r_1} \right]^{1/r_1}$ indica a dispersão interna à uma partição i e $d_{ij,t} = \|\mathbf{w}_i - \mathbf{w}_j\|_{r_2}$ indica a separação entre partições dada pela distância de Minkowski. Por fim, para o índice de Calinski-Harabasz $\mathbf{S}_B = \sum_{i=1}^K N_i (\mathbf{w}_i - \mathbf{m})(\mathbf{w}_i - \mathbf{m})^T$ indica a dispersão entre grupos e

$\mathbf{S}_W = \sum_{i=1}^K N_i \mathbf{C}_i$ indica a dispersão interna ao grupo. Vale ainda ressaltar que N indica o número de amostras no conjunto de dados e K o número escolhido para a quantidade de agrupamentos. Em mente de tais definições, os índices calculados são apresentados abaixo

Tabela 1: Tabela de Índices para K-Means sequencial.

K	Dunn	Davies-Bouldin	Calinski-Harabasz
2	0.012077	0.79934	591.1587
3	0.0048257	0.89444	504.847
4	0.005162	0.90558	500.837
5	0.0060383	0.89991	528.5272
6	0.0068242	0.86367	558.6935
7	0.0058744	0.81966	591.1628
8	0.0043136	0.78708	615.8837
9	0.0041443	0.77143	628.3026
10	0.0042644	0.77543	623.351
11	0.0038419	0.76322	628.8014
12	0.0035803	0.75345	630.3934
13	0.0033439	0.73317	633.8261
14	0.0032612	0.70736	645.577
15	0.0031304	0.70205	634.938
16	0.0032568	0.69308	627.3569
17	0.0031896	0.68632	621.8538
18	0.0030558	0.692	610.9051
19	0.003099	0.68771	597.4508
20	0.003196	0.6951	587.1598

Tabela 2: Tabela de Índices para K Means em batch.

K	Dunn	Davies-Bouldin	Calinski-Harabasz
2	0.014182	0.81182	569.8278
3	0.0046508	0.9441	486.0957
4	0.0044554	0.98955	466.7917
5	0.0049625	0.97861	472.4607
6	0.0044633	0.97477	485.054
7	0.0040531	0.97509	509.4936
8	0.0038761	0.94252	524.7318
9	0.0037034	0.90832	542.6308
10	0.0035293	0.89519	553.9311
11	0.0034286	0.87275	553.6746
12	0.0033552	0.85905	552.1686
13	0.0033632	0.84977	557.5972
14	0.0033259	0.82024	560.8382
15	0.0033082	0.81705	557.4471
16	0.0032715	0.81004	552.7867
17	0.0033662	0.79918	554.9703
18	0.0033463	0.78006	554.5958
19	0.0032699	0.78348	541.3209
20	0.0031833	0.77777	543.1308

De acordo com as tabelas 1 e 2 e com figuras 1 e 2 é possível intuir que o número de clusters ótimo para o conjunto de dados será algo no intervalo 13-15. Ademais, é interessante notar que embora tenha existido uma divergência entre os índices todos apontaram, para ambos os algoritmos, um valor ótimo existente entre 13-15. Isso poderia ser explicado pelas diferentes definições de coesão interna ou dispersão externa que cada um dos índices utiliza em sua definição. Além disso, é interessante ainda citar que para o índice de Davies-Bouldin escolhi utilizar os valores $r_1 = r_2 = 2$, o que também poderia vir a ter algum impacto na divergência entre os resultados apontados por cada índice. Apesar de tudo, é interessante notar que apesar de tais

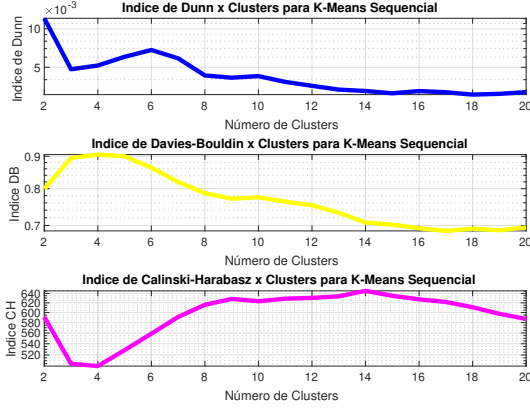


Figura 1: K-Means Sequencial: Índices x Número de Clusters

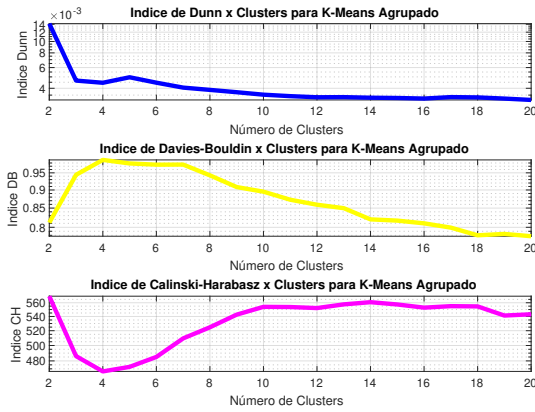


Figura 2: K-Means em Batch: Índices x Número de Clusters

divergências todos os índices apontam valores aproximadas para o número de clusterização.

Em continuidade, chamo atenção ao fato de que o desempenho para um número de clusters igual a 10 não difere tanto assim dos valores ótimos acima citado, o que pode ser visualizado mais facilmente nas Figuras 1 e 2. Desse modo, para facilitar tanto a escrita quanto a correção do relatório decidi por adotar o valor sub-ótimo $K_{opt} = 10$. Posso também justificar a adoção desse valor sub-ótimo levando-se em consideração o recorrente compromisso entre desempenho e complexidade presente na implementação de qualquer sistema prática teleinformática. Evidentemente é benéfico para o desempenho final do K-Means o uso do número ideal de clusters. Contudo, tal valor poderia ser considerado desnecessariamente custoso computacionalmente para aplicações que venham a utilizar o algoritmo, visto que processamento computacional complexo pode acarretar em maiores gastos financeiros para a implementação satisfatória de certas aplicações. Isso pode ocorrer uma vez que o tempo de processa-

mento dos dados pode afetar diretamente o desempenho final de aplicações que venham a depender de um tempo de resposta com baixíssima latência.

- As posições dos protótipos são apresentadas nas tabelas abaixo, mas antes disso também disponho a visualização dos gráficos para a SSD por iteração para os dois métodos utilizados. É possível verificar que na Figura 3 que as posições dos protótipos já não apresentam atualizações substanciais após a iteração 50, enquanto na Figura 4 é possível conferir que a convergência do algoritmo ocorreu de forma satisfatória na décima iteração. Por fim, as posições dos protótipos de cada classe estão dispostas nas Tabelas 3 e 4 para o K-Means sequencial e para o K-Means em Batch, respectivamente.

No mais, é interessante notar que os protótipos não convergiram para os mesmos valores nessa simulação. Isso pode ser explicado por alguns fatores. O primeiro pelo próprio conjunto de dados ter uma quantidade limitada de amostras e isso poderia de algum modo estar prejudicando o processo de convergência para os centróides dos agrupamentos. Em segundo lugar as tabelas foram gerados para um valor que não é ótimo devido as razões alencadas anteriormente. Ao realizar alguns testes um pouco menos complexos foi possível perceber que para os valores supostamente ótimos alguns dos protótipos de fato convergem para o mesmo centróide o que indicaria a equivalência entre os dois algoritmos. Entretanto, não vi em nenhum caso uma convergência perfeita de todos os protótipos.

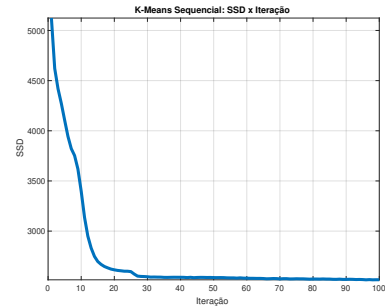


Figura 3: K-Means Sequencial: SSD x Iteração

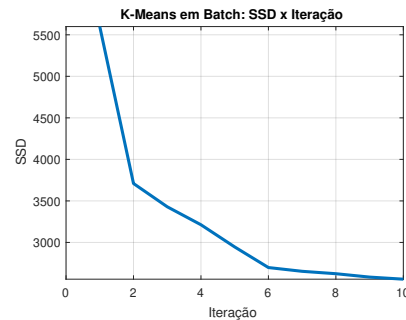


Figura 4: K-Means em Batch: SSD x Iteração

Tabela 3: Protótipos K-Means sequencial

	Par1	Par2	Par3	Par4	Par5	Par6
w_1	0.038027	-0.20488	0.19666	-0.90615	-0.88964	-1.0445
w_2	0.14455	-0.57053	-0.25237	1.2419	1.2211	0.93905
w_3	0.044981	0.27469	0.34004	-1.2066	-1.2436	0.9643
w_4	0.11521	-0.19088	-0.20753	0.14463	0.16901	0.95297
w_5	-7.3905	0.34211	0.13765	-0.3268	-0.22316	-0.64196
w_6	-0.32101	0.47864	9.8847	-1.5422	-1.563	-1.015
w_7	0.14455	0.2514	-0.25473	1.2419	1.2211	0.95024
w_8	0.085577	3.6663	0.062673	-0.071102	-0.012304	-0.61682
w_9	0.14231	-0.19657	-0.25229	0.85514	0.85651	-1.0502
w_{10}	0.13172	-0.15136	-0.11933	-0.60506	-0.6127	0.96194

Tabela 4: Protótipos K-Means em Batch

	Par1	Par2	Par3	Par4	Par5	Par6
w_1	0.13154	-0.14775	-0.17621	-0.58125	-0.56269	-1.0418
w_2	0.14455	-0.20307	-0.25002	1.2027	1.1927	0.93988
w_3	0.14455	-0.58363	-0.27293	1.2419	1.2211	-1.0673
w_4	0.14455	-0.22125	-0.21833	0.14306	0.19243	-1.0488
w_5	0.097114	-0.056044	-0.0070867	-0.61174	-0.62712	0.96076
w_6	-0.52637	0.3894	7.268	-1.5154	-1.5078	-0.92962
w_7	-0.038091	-0.19577	0.41873	-1.2684	-1.2672	-1.0498
w_8	-7.4505	0.2247	-0.085986	-0.29551	-0.20663	-0.61893
w_9	0.08938	3.5571	0.098054	-0.24377	-0.19104	-0.33794
w_{10}	0.12173	0.24357	-0.2534	1.2325	1.2192	-1.03

- As estatísticas de agrupamento para $K_{\text{opt}} = 10$ são apresentadas abaixo nas Tabelas 5-14 e 15-24 para o K-Means sequencial e para o K-Means em Batch, respectivamente.

3 Conclusão

Neste trabalho, foi possível conferir matematicamente a equivalência entre os algoritmos de clusterização K-Means sequencial e Batch. Além disso, também foi possível analisar o uso de diferentes métricas qualitativas de desempenho para algoritmos de clusterização. Não menos importante, foram também estudados alguns conceitos relevantes para o tópico como a análise de convergência dos algoritmos, procedimento para determinação do hiperparâmetro que define o número ótimo para a clusterização e definição de coesão interna e dispersão externa.

Tabela 5: K-Means Sequencial: Estatísticas do Agrupamento 1

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	0.046803	0.38441
Par2	-1.1475	2.0465	-0.30904	-0.16933	0.44261
Par3	-0.2681	4.8171	-0.12917	0.18492	0.81869
Par4	-1.8764	-0.10658	-0.8921	-0.91604	0.42783
Par5	-1.9302	0.24429	-0.88064	-0.89971	0.45994
Par6	-1.0968	-0.88607	-1.0705	-1.0447	0.058929

Tabela 6: K-Means Sequencial: Estatísticas do Agrupamento 2

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	0.14455	0.14455	0.14455	0.14455	5.5674e-17
Par2	-0.87453	-0.19204	-0.60154	-0.572	0.10521
Par3	-0.29141	0.098996	-0.27615	-0.25196	0.068196
Par4	1.2419	1.2419	1.2419	1.2419	8.9079e-16
Par5	1.2211	1.2211	1.2211	1.2211	2.227e-16
Par6	0.65514	1.0929	0.9557	0.93448	0.10497

Tabela 7: K-Means Sequencial: Estatísticas do Agrupamento 3

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	0.03566	0.47207
Par2	-0.9352	2.7708	0.17521	0.27254	0.64491
Par3	-0.2207	4.9159	0.098122	0.33993	0.70311
Par4	-1.7082	-0.60327	-1.1989	-1.199	0.20959
Par5	-1.778	-0.46002	-1.2528	-1.2364	0.23093
Par6	0.65515	1.0721	0.95451	0.96448	0.035181

Tabela 8: K-Means Sequencial: Estatísticas do Agrupamento 4

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	0.12103	0.28218
Par2	-1.011	1.4459	-0.32854	-0.19262	0.40783
Par3	-0.26979	1.1724	-0.25218	-0.20554	0.14744
Par4	-0.34311	0.47459	0.16106	0.13406	0.14581
Par5	-0.26927	0.89758	0.10724	0.16319	0.17686
Par6	0.65523	1.0782	0.95667	0.9541	0.063416

Tabela 9: K-Means Sequencial: Estatísticas do Agrupamento 5

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-16.7862	-4.0881	-5.499	-7.3656	3.6588
Par2	-1.1475	2.8928	0.12646	0.3407	1.1247
Par3	-0.28146	5.082	-0.20931	0.13871	1.1299
Par4	-1.2303	1.2419	-0.34311	-0.32941	0.54564
Par5	-1.1854	1.2211	-0.41232	-0.22526	0.62043
Par6	-1.098	0.96083	-1.0701	-0.63855	0.83018

Tabela 10: K-Means Sequencial: Estatísticas do Agrupamento 6

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	-0.35248	1.1054
Par2	-0.49001	5.823	-0.30254	0.53242	2.0225
Par3	5.8078	24.207	8.0246	10.2409	5.7665
Par4	-1.802	-0.79488	-1.695	-1.5404	0.34011
Par5	-1.8987	-0.8556	-1.6434	-1.5575	0.31263
Par6	-1.0822	-0.96832	-1.0706	-1.0504	0.046764

Tabela 11: K-Means Sequencial: Estatísticas do Agrupamento 7

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	0.14455	0.14455	0.14455	0.14455	8.3548e-17
Par2	-0.05554	2.6744	0.21746	0.25868	0.5072
Par3	-0.28898	0.027129	-0.27725	-0.25542	0.057766
Par4	1.2419	1.2419	1.2419	1.2419	6.6838e-16
Par5	1.2211	1.2211	1.2211	1.2211	2.2279e-16
Par6	0.65512	1.0699	0.95802	0.94508	0.079966

Tabela 12: K-Means Sequencial: Estatísticas do Agrupamento 8

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-1.7366	0.14455	0.14455	0.081207	0.32769
Par2	1.9464	18.6903	2.6744	3.7121	2.9773
Par3	-0.29101	2.081	-0.14665	0.070369	0.54163
Par4	-1.4121	1.2419	-0.2737	-0.07454	0.98536
Par5	-1.514	1.2211	-0.18508	-0.011996	0.95928
Par6	-1.0892	1.0281	-0.94316	-0.60233	0.79562

Tabela 13: K-Means Sequencial: Estatísticas do Agrupamento 9

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-2.4797	0.14455	0.14455	0.13722	0.1387
Par2	-0.87453	1.5824	-0.32854	-0.21259	0.43662
Par3	-0.2929	1.1157	-0.2793	-0.25093	0.11379
Par4	-0.20429	1.2419	1.2419	0.8711	0.51158
Par5	0.10724	1.2211	1.2211	0.8706	0.49207
Par6	-1.1002	-0.886	-1.0716	-1.0496	0.060471

Tabela 14: K-Means Sequencial: Estatísticas do Agrupamento 10

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-2.2741	0.14455	0.14455	0.13412	0.15879
Par2	-0.97691	1.4186	-0.21544	-0.14494	0.31921
Par3	-0.25389	0.84938	-0.1842	-0.12896	0.16736
Par4	-1.14	-0.20429	-0.64165	-0.59851	0.25107
Par5	-1.2059	0.42978	-0.61419	-0.60961	0.27222
Par6	0.6552	1.0721	0.9557	0.96063	0.048802

Tabela 15: K-Means em Batch: Estatísticas do Agrupamento 1

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-2.2741	0.14455	0.14455	0.13154	0.17735
Par2	-0.69253	1.1275	-0.25346	-0.14775	0.40009
Par3	-0.2681	0.68662	-0.20527	-0.17621	0.1168
Par4	-1.2819	-0.20429	-0.66104	-0.58125	0.22462
Par5	-1.0666	-0.065171	-0.61419	-0.56269	0.23616
Par6	-1.093	-0.88607	-1.0705	-1.0418	0.063934

Tabela 16: K-Means em Batch: Estatísticas do Agrupamento 2

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	0.14455	0.14455	0.14455	0.14455	8.3392e-17
Par2	-0.96553	2.4014	-0.32854	-0.20307	0.49443
Par3	-0.29141	0.098996	-0.27615	-0.25002	0.066356
Par4	0.16106	1.2419	1.2419	1.2027	0.18861
Par5	0.2561	1.2211	1.2211	1.1927	0.14461
Par6	0.65512	1.0929	0.95691	0.93988	0.094282

Tabela 17: K-Means em Batch: Estatísticas do Agrupamento 3

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	0.14455	0.14455	0.14455	0.14455	5.574e-17
Par2	-0.87453	-0.32854	-0.60154	-0.58363	0.084072
Par3	-0.29199	-0.013795	-0.28544	-0.27293	0.040734
Par4	1.2419	1.2419	1.2419	1.2419	0
Par5	1.2211	1.2211	1.2211	1.2211	2.2296e-16
Par6	-1.1002	-0.91896	-1.0804	-1.0673	0.033919

Tabela 18: K-Means em Batch: Estatísticas do Agrupamento 4

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	0.14455	0.14455	0.14455	0.14455	5.5723e-17
Par2	-0.60154	0.89995	-0.32854	-0.22125	0.36663
Par3	-0.28099	0.63978	-0.26191	-0.21833	0.13033
Par4	-0.38172	0.47459	0.16106	0.14306	0.12477
Par5	-0.1141	0.91094	0.10724	0.19243	0.155
Par6	-1.093	-0.88607	-1.0715	-1.0488	0.059086

Tabela 19: K-Means em Batch: Estatísticas do Agrupamento 5

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	0.097114	0.32988
Par2	-1.011	2.2844	-0.12379	-0.056044	0.41588
Par3	-0.26979	3.356	-0.16133	-0.0070867	0.43893
Par4	-1.7082	0.30551	-0.66104	-0.61174	0.53087
Par5	-1.778	0.43035	-0.69687	-0.62712	0.5518
Par6	0.6552	1.0721	0.95563	0.96076	0.046156

Tabela 20: K-Means em Batch: Estatísticas do Agrupamento 6

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-5.499	0.14455	0.14455	-0.52637	1.4759
Par2	-0.49001	5.823	-0.27043	0.3894	1.5766
Par3	3.7082	24.207	5.4449	7.268	5.0142
Par4	-1.802	-0.79488	-1.5408	-1.5154	0.26954
Par5	-1.8987	-0.63524	-1.598	-1.5078	0.32257
Par6	-1.093	0.94213	-1.0702	-0.92962	0.4703

Tabela 21: K-Means em Batch: Estatísticas do Agrupamento 7

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-3.2416	0.14455	0.14455	-0.038091	0.51064
Par2	-1.1475	2.0465	-0.36266	-0.19577	0.49013
Par3	-0.23588	3.3718	0.11524	0.41873	0.78365
Par4	-1.8764	-0.72812	-1.2508	-1.2684	0.23703
Par5	-1.9302	-0.38038	-1.2335	-1.2672	0.27146
Par6	-1.0968	-0.88649	-1.0703	-1.0498	0.051696

Tabela 22: K-Means em Batch: Estatísticas do Agrupamento 8

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-16.7862	-4.0881	-5.499	-7.4505	3.7217
Par2	-1.1475	2.6744	0.069584	0.2247	1.0005
Par3	-0.28146	1.0845	-0.21551	-0.085986	0.3477
Par4	-1.2303	1.2419	-0.34311	-0.29551	0.5331
Par5	-1.1854	1.2211	-0.41232	-0.20663	0.62841
Par6	-1.098	0.96083	-1.0701	-0.61893	0.84424

Tabela 23: K-Means em Batch: Estatísticas do Agrupamento 9

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-1.7366	0.14455	0.14455	0.08938	0.3062
Par2	1.9464	18.6903	2.6744	3.5571	2.8057
Par3	-0.29101	2.081	-0.098319	0.098054	0.51199
Par4	-1.4865	1.2419	-0.53899	-0.24377	1.0203
Par5	-1.6542	1.2211	-0.40615	-0.19104	1.0113
Par6	-1.0832	1.033	-0.94302	-0.33794	0.92908

Tabela 24: K-Means em Batch: Estatísticas do Agrupamento 10

	Mínimo	Máximo	Mediana	Média	Desvio
Par1	-2.4797	0.14455	0.14455	0.12173	0.24471
Par2	-0.05554	2.4014	-0.05554	0.24357	0.4503
Par3	-0.2929	1.1157	-0.28674	-0.2534	0.14374
Par4	0.16106	1.2419	1.2419	1.2325	0.10079
Par5	0.99512	1.2211	1.2211	1.2192	0.021076
Par6	-1.0953	-0.886	-1.0714	-1.03	0.077359