

Neste trabalho escolhi como características do algoritmo de ES mutar de forma não correlacionada com apenas um sigma por cromossomo; o método de seleção foi realizado de forma (30+210), sendo que cada pai produz 6 filhos por mutação e 1 por combinação; o método de combinação feito foi de escolha de 2 pais fixos e o filho produzido é a média de ambos os pais.

O algoritmo de DE foi feito conforme com a descrição do artigo FSDE-Forced Strategy Differential Evolution used for data clustering, com a única diferença que a taxa Cr para crossover é fixa e igual 0,66. Com isso, se um número escolhido de forma aleatória entre 0 e 1, se for menor que 0,66, o filho recebe o gene do pai mutado, caso contrário recebe o gene do pai original.

Este trabalho foi realizado em duas etapas chamadas treino e teste. Na fase de treino foi utilizado 75% de cada base de dados do trabalho para treinar os clusters implementados neste trabalho e na fase de teste foi utilizado 25% da base de dados para testar os algoritmos implementados neste trabalho.

Uma das métricas utilizadas nesse trabalho foi a soma dos erros quadráticos (sum of squared errors – SSE) dado pela fórmula abaixo:

$$SSE = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - m_i\|^2$$

onde  $m_i$  é o centróide do cluster  $C_i$ , e  $x_j$  é um vetor de característica pertencente ao cluster  $C_i$ .

Abaixo é apresentado uma tabela apresentando a média seguida da variância dos valores SSE dos algoritmos DE, ES e suas respectivas versões hibridizadas com K-means. Esses valores foram obtidos após 10 execuções de cada algoritmo.

	ES	DE	ES-K-means	DE-K-means
Treino Iris	99.11 ± 17.28	71.25 ± 5.76	61.30 ± 0.0	61.30 ± 0.0
Teste Iris	97.67 ± 33.09	23.89 ± 2.98	134.76 ± 49.782	19.61 ± 0.0
Treino Wine	1956905.31 ± 24183.70	1989276.14 ± 11418.27	1911759.59 ± 0.0	1911759.59 ± 0.0

Teste Wine	1281985.77 ± 637278.37	601713.76 ± 129236.96	3541496.20 ± 3651739.17	483597.10 ± 0.0
Treino Cancer	29265.41 ± 4040.19	14591.16 ± 7.71	14578.51 ± 0.0	14578.51 ± 0.0
Teste Cancer	23261.26 ± 7543.09	4870.82 ± 14.24	24904.36 ± 4553.12	4851.56 ± 0.0

Abaixo é apresentado uma tabela apresentando a média seguida da variância dos valores de acurácia em porcentagem dos algoritmos DE, ES e suas respectivas versões hibridizadas com K-means.

Esses valores foram obtidos após 10 execuções de cada algoritmo.

	ES	DE	ES-K-means	DE-K-means
Treino Iris	87.589 ± 2.750	88.125 ± 2.076	90.178 ± 0.0	90.178 ± 0.0
Teste Iris	77.368 ± 16.734	87.368 ± 3.867	66.578 ± 18.572	89.473 ± 0.0
Treino Wine	59.248 ± 4.729	59.699 ± 5.112	63.157 ± 0.0	63.157 ± 0.0
Teste Wine	68.666 ± 10.108	74.666 ± 13.813	56.222 ± 25.435	84.444 ± 0.0
Treino Cancer	92.089 ± 2.314	95.703 ± 0.0	95.703 ± 0.0	95.703 ± 0.0
Teste Cancer	74.260 ± 26.144	98.816 ± 0.0	60.473 ± 30.266	98.816 ± 0.0

Uma outra métrica utilizada nesse foi a distância entre os clusters (inter cluster distance – ICE) dado pela fórmula abaixo:

$$ICE = \min (c_i - c_j)^2$$

onde  $c_i$  e  $c_j$  são centróides de um cluster e  $i \neq j$ .

Abaixo é apresentado uma tabela apresentando a média seguida da variância dos valores ICE dos algoritmos DE, ES e suas respectivas versões hibridizadas com K-means.

Esses valores foram obtidos após 10 execuções de cada algoritmo.

	ES	DE	ES-K-means	DE-K-means
Treino Iris	$4.73 \pm 2.27$	$2.78 \pm 0.71$	$4.22 \pm 1.90$	$3.35 \pm 0.0$
Teste Iris	$3.01 \pm 1.39$	$3.00 \pm 0.94$	$3.86 \pm 2.11$	$3.35 \pm 0.0$
Treino Wine	$70348.87 \pm 58323.67$	$98518.13 \pm 38084.47$	$49005.94 \pm 51153.16$	$68757.61 \pm 0.0$
Teste Wine	$70975.34 \pm 55411.53$	$98518.13 \pm 38084.47$	$57825.16 \pm 52050.93$	$68757.61 \pm 0.0$
Treino Cancer	$125.22 \pm 28.76$	$182.26 \pm 1.22$	$102.37 \pm 27.23$	$181.18 \pm 0.0$
Teste Cancer	$125.22 \pm 28.76$	$182.26 \pm 1.22$	$102.37 \pm 27.23$	$181.18 \pm 0.0$

Podemos concluir que os algoritmos implementados neste trabalho, apresentaram bons resultados para classificação para os datasets da Iris e do Cancer, já que os dados nesses datasets são lineares em sua maioria. Já no dataset Wine, os algoritmos apresentaram resultados ruins para classificação, pois os dados neste dataset não são lineares. As versões hibridizadas com uso do K-means tanto no algoritmo DE e do ES, apresentaram melhores resultados para clusterização no SSE do que suas versões normais, já que o valor do SSE é menor nas suas versões hibridizadas do que nas suas versões normais. Por fim, o melhor algoritmo implementado neste trabalho foi o DE hibridizado com K-means, devido apresentar na média apresentar os melhores resultados e por ser o mais estável.