African Vultures Optimization (AVO) for Feature Selection

Analysis of Algorithms Project by

SULTAN | 56189 | BS/DS-4_1

Instructor : Sir Usman Shareef

The Problem of High-Dimensional Data in Machine Learning

Performance Degradation

High-dimensional datasets can lead to degraded machine learning model performance due to noisy or irrelevant data.

Inefficient Training

Irrelevant or redundant features increase computational costs and training time without improving results.

Need for Feature Selection

Efficient, fast, and accurate feature selection methods are critical for optimal model accuracy and scalability.

The Solution: African Vultures Optimization Algorithm

Nature-Inspired Algorithm

AVO mimics vultures' energy-efficient scavenging strategies for optimization.

Binary Feature Selection

Selects subsets of features represented as binary vectors where 1 means selected, 0 ignored.

Balanced Search

Maintains a balance between exploration of new solutions and exploitation of promising feature subsets.

Dataset Overview: Breast Cancer Wisconsin Dataset

Dataset Characteristics

- 569 samples with 30 numeric features each
- Binary target: benign vs malignant tumors
- Widely used benchmark for classification and feature selection

Preprocessing

Data was split into training and testing sets to evaluate algorithm performance and prevent overfitting.

This setup allows reliable comparison of feature selection effectiveness.

Code Design & Fitness Function

Binary Encoding

Features represented as binary vectors indicating whether to include each feature.

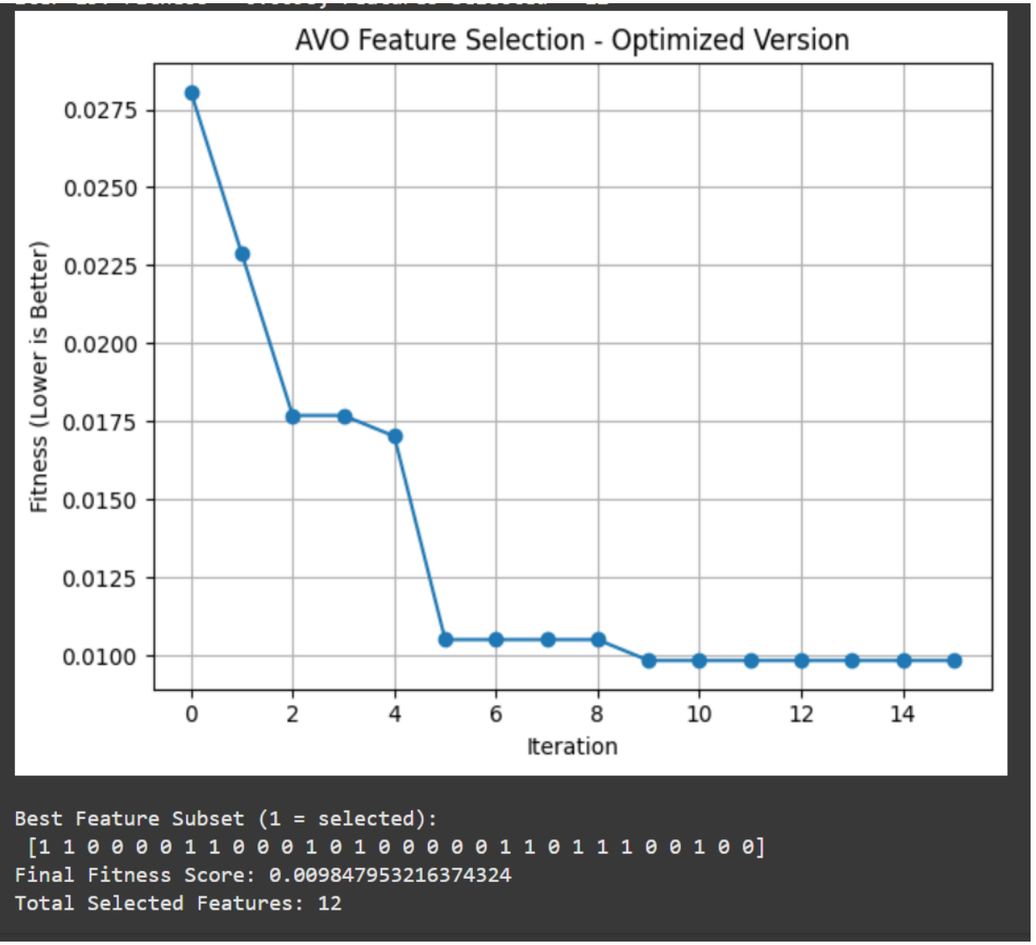
Fitness Function Definition

The function optimizes for classification accuracy penalized by the number of selected features:
(1 - Accuracy) + 0.01 × Feature Ratio

Logistic Regression serves as the evaluation model.

Convergence Monitoring

Algorithm progress tracked over 15 iterations to ensure effective feature subset selection.



Results: Convergence & Performance

1

Feature Reduction

AVO effectively reduced the number of features while maintaining classification accuracy.

2

Fitness Improvement

Fitness score consistently improved throughout the iterations, demonstrating algorithm stability and efficiency.

3

Small & Effective Subset

The final subset selected was compact yet highly effective for the classification task.

Comparative Advantages and Real-World Impact

Algorithmic Benefits

- Superior ability to escape local optima
- Faster convergence rates
- Requires fewer fitness evaluations

Use Case Impact

- Accelerates medical diagnosis by reducing feature dimensionality
- Enables compact models for embedded AI applications
- Energy-efficient, suitable for resource-constrained environments

Conclusion & Future Directions

Effectiveness of AVO

Demonstrated strong potential as a feature selection method by merging swarm intelligence with practical constraints.

Future Work

Plan to extend AVO applications to diverse domains including Internet of Things, finance, and text analytics for broader impact.

Engagement

Open for questions and collaborative discussions on further enhancement and deployment.

THANK YOU

FOR

YOUR TIME