Support Vector Machine (SVM) Documentation

1. Introduction

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data into distinct classes.

2. Objective

- Develop an interactive web application using Streamlit for SVM classification.
- Provide users with the ability to upload a dataset, select features, train the model, and evaluate its performance.

3. Dataset Description

- Dataset: **A suitable dataset**
- Features: Describe the input variables
- Target: Describe the variable to predict
- Number of records and attributes

4. Implementation Details

- **Frontend**: Developed using Streamlit for user interaction.
- **Backend**: SVM implemented using `scikit-learn`.
- **Steps**:
  1. Upload the dataset in CSV format.
  2. Select features and target variables.
  3. Choose kernel type (Linear, Polynomial, RBF, or Sigmoid).
  4. Train the SVM model.

5. Evaluate the model using accuracy, confusion matrix, and classification report.

5. Results and Analysis

- Provide model accuracy and other evaluation metrics.
- Visualize the decision boundary for 2D datasets.
- Discuss the impact of kernel choice on model performance.

6. Challenges and Solutions

- Managed large datasets using appropriate kernel functions.
- Optimized hyperparameters using GridSearchCV.
- Addressed class imbalances using preprocessing techniques.

7. Conclusion

Support Vector Machines are effective for both linear and non-linear classification tasks. The interactive application allows users to explore different kernel functions and observe their effects on model performance.

8. References

- Scikit-learn Documentation
- Streamlit Documentation
- Dataset Source