

Breast Cancer Classification using (SVM)

Internship Task 7

1. Abstract

This project implements Support Vector Machines (SVM) for binary classification using the Breast Cancer Wisconsin dataset. Both linear and non-linear (RBF) kernels are applied, followed by hyperparameter tuning, decision boundary visualization, and cross-validation to achieve high predictive accuracy and model reliability.

2. Problem Statement

To develop a robust and accurate machine learning model capable of classifying breast tumors as benign or malignant using Support Vector Machines with optimized hyperparameters.

3. Dataset Description

The Breast Cancer Wisconsin dataset consists of clinical features extracted from digitized images of breast mass biopsies. It contains 569 samples with 30 numerical features. The target variable indicates whether the tumor is benign or malignant.

4. Methodology

- Data Loading and Exploration
- Feature Scaling using Standardization
- Train-Test Split
- Linear Kernel SVM Training
- RBF Kernel SVM Training
- Hyperparameter Tuning using GridSearchCV
- Decision Boundary Visualization using PCA
- Cross-Validation Evaluation

5. Model Training

Two SVM models were trained using linear and RBF kernels. The RBF kernel provided superior performance due to its ability to capture non-linear decision boundaries. Feature scaling significantly improved convergence and classification accuracy.

6. Hyperparameter Tuning

GridSearchCV was used to tune the hyperparameters C and gamma for the RBF kernel. This optimization significantly improved the model's generalization and predictive performance.

7. Decision Boundary Visualization

Principal Component Analysis (PCA) was applied to reduce dimensionality for 2D visualization. The resulting decision boundary clearly demonstrates the separation between malignant and benign tumor classes.

8. Model Evaluation

The optimized SVM model achieved high classification accuracy. Cross-validation confirmed the robustness and stability of the model across multiple data splits.

9. Conclusion

This project successfully demonstrates the application of Support Vector Machines for medical diagnosis. The integration of feature scaling, hyperparameter tuning, visualization, and cross-validation ensures a highly accurate, reliable, and interpretable classification system suitable for clinical decision support.