Name - Jatin Mahawar
Roll.No. - 22CS30032

Machine Learning Assignment 3 Part A Report

SVM with Different Kernels: Comparison and Results

Several SVM models with different kernel types were trained and evaluated. The purpose of using various kernels (such as linear, polynomial, and RBF) was to understand which kernel best captures the relationships in the data and achieves optimal performance.

1. Overview of Kernels and Their Suitability

- **Linear Kernel**:
    - Suitable for linearly separable data or when a linear decision boundary is expected to suffice.
    - Fastest among SVM kernels because it computes a straightforward linear boundary.
- **Polynomial Kernel**:
    - Useful when interactions between features are significant and a non-linear decision boundary is required.
    - Allows the model to fit complex, polynomial relationships but can be computationally intensive, especially as the degree increases.
- **RBF (Radial Basis Function) Kernel**:
    - Ideal for capturing non-linear relationships with smooth decision boundaries, as it considers the distance of points from each other.
    - Flexible and can adapt to more complex structures but generally requires more computational power due to non-linear transformations.

2. Training and Cross-Validation

- Each SVM model with a different kernel was trained using cross-validation to ensure robust performance measurement.
- Cross-validation splits helped to mitigate overfitting and provided an average accuracy across folds, giving a reliable indicator of each kernel's generalizability.

3. Evaluation Metrics

- The main metrics for comparing the SVM models with different kernels included:
    - **Accuracy**: The percentage of correct predictions, giving a high-level view of each model's effectiveness.
    - **Precision and Recall**: These were particularly useful for understanding how well each model handled positive and negative predictions.
    - **F1 Score**: This balanced metric provided a single score reflecting both precision and recall.
    - **AUC (Area Under Curve)**: Critical for binary classification, the AUC measured each model's ability to separate classes, highlighting kernel efficacy in distinguishing between positive and negative classes.

4. Results and Comparison of Kernels

- **Polynomial Kernel**:
  - **Accuracy:** 0.56
  - **Precision:** 0.56
  - **Recall:** 1
  - **F1-score:** 0.72
  - **AUC:** 0.52
  - **Best parameters**: {'kernel': 'poly', 'degree': 3, 'C': 0.1}
  - **Best score**: 0.5718253165126268
  - **Training time**: 4.6389806270599365 seconds
  - **Prediction time**: 0.025223255157470703 seconds
- **RBF Kernel**:
  - **Accuracy:** 0.56
  - **Precision:** 0.56
  - **Recall:** 1
  - **F1-score:** 0.72
  - **AUC:** 0.53
  - **Best parameters:** {'kernel': 'rbf', 'gamma': 0.1, 'C': 10.0}
  - **Best score:** 0.5704535744001439
  - **Training time:** 5.430238485336304 seconds
  - **Prediction time:** 0.07356119155883789 seconds
- **Sigmoid Kernel**:
  - **Accuracy:** 0.56
  - **Precision:** 0.56
  - **Recall:** 1
  - **F1-score:** 0.72
  - **AUC:** 0.47
  - **Best parameters**: {'kernel': 'sigmoid', 'coef0': 0, 'C': 0.1}
  - **Best score**: 0.5704535744001439
  - **Training time**: 5.244624137878418 seconds
  - **Prediction time**: 0.04303693771362305 seconds

5. Observations and Final Kernel Choice

- **Best Performance**: The **RBF kernel** outperformed the others in accuracy and AUC, making it the best choice if the goal is to maximize classification performance on this dataset.
- **Balanced Choice**: The **polynomial kernel** was a middle ground, offering better accuracy than the linear kernel but with less computational cost than RBF, potentially suitable if some non-linear relationships are suspected but full RBF complexity is unnecessary.