CS60050_Machine Learning_Programming Assignment_3

Part A: Support Vector Machines (SVMs) and Kernel Methods - HIGGS Dataset

Chandransh Singh (22CS30017)

1. **Linear Kernel**
   - Training Time: 0.011618 seconds
   - Accuracy: 0.55
   - Classification Report:
     - Precision: 0.62 (for class 0), 0.55 (for class 1)
     - Recall: 0.14 (for class 0), 0.92 (for class 1)
     - F1-Score: 0.23 (for class 0), 0.69 (for class 1)
   - AUC: 0.56

2. **Polynomial Kernel**
   - Training Time: 0.0206 seconds
   - Accuracy: 0.61
   - Classification Report:
     - Precision: 0.72 (for class 0), 0.58 (for class 1)
     - Recall: 0.28 (for class 0), 0.91 (for class 1)
     - F1-Score: 0.40 (for class 0), 0.71 (for class 1)
   - AUC: 0.592

3. **RBF Kernel**
   - Training Time: 0.0119 seconds
   - Accuracy: 0.65
   - Classification Report:
     - Precision: 0.65 (for class 0), 0.64 (for class 1)
     - Recall: 0.54 (for class 0), 0.74 (for class 1)
     - F1-Score: 0.59 (for class 0), 0.69 (for class 1)
   - AUC: 0.639

4. **Custom Kernel**
   - Training Time: 0.0048 seconds
   - Accuracy: 0.59
   - Classification Report:
     - Precision: 0.57 (for class 0), 0.61 (for class 1)
     - Recall: 0.57 (for class 0), 0.61 (for class 1)

- F1-Score: 0.57 (for class 0), 0.61 (for class 1)
  - AUC: 0.589

Performance Comparison

Accuracy:

- Best: RBF Kernel (0.65)
- Moderate: Polynomial Kernel (0.61)
- Good but lower: Custom Kernel (0.59)
- Lowest: Linear Kernel (0.55)

Precision and Recall:

- RBF Kernel demonstrates a balanced precision and recall for both classes, indicating a reliable model performance.
- Polynomial Kernel has higher precision for class 0 but lower recall, suggesting it is better at avoiding false positives but missing some true positives.
- Linear Kernel shows very low recall for class 0, which means it is struggling to identify that class.
- Custom Kernel maintains balanced precision and recall, but at lower values than RBF.

F1-Score:

- The RBF Kernel again performs best overall, particularly in identifying class 1 correctly while still maintaining a decent performance for class 0.
- The Polynomial Kernel shows the highest F1-score for class 1, while the Linear Kernel has the lowest F1-score for class 0.

AUC Scores:

- Best: RBF Kernel (0.639)
- Moderate: Polynomial Kernel (0.592)
- Good: Custom Kernel (0.589)
- Lowest: Linear Kernel (0.56)

Training and Prediction Time:

- The Custom Kernel is the fastest in terms of training time (0.0048 seconds), followed closely by the RBF Kernel (0.0119 seconds) and Linear Kernel (0.011618 seconds).
- Polynomial Kernel takes longer to train (0.0206 seconds), suggesting more complexity in its calculations.

Insights and Recommendations

Most Suitable Kernel: The `RBF Kernel` stands out as the best option for the HIGGS dataset, as it offers the highest accuracy (0.65), balanced precision and recall, and the best AUC score (0.639). It also maintains efficient training time, making it a robust choice for practical applications.

Use Cases for Other Kernels:

- Polynomial Kernel may be appropriate if you require a slightly more complex decision boundary and can afford the longer training time.
- Linear Kernel might be useful for very large datasets where computational efficiency is critical, despite its poorer performance on this specific dataset.
- Custom Kernel could serve as an alternative if specific domain knowledge justifies its formulation, but it currently underperforms compared to the others.

Overall, the choice of kernel should balance performance metrics with computational efficiency, with the RBF Kernel being the most favorable for this dataset.