Portfolio Assessment Submission

Student Information

Name: Arnob Ghosh

Student Number: 103494114

Studio Class: Monday 4:30 – 6:30 PM

Summary Table of Studio 3: Activity 6

LINK: Week 3

I concentrated on creating and assessing a Support Vector Machine (SVM) model using various setups in Studio 3, Activity 6. Finding out how different adjustments, like feature selection, dimensionality reduction using PCA, and hyperparameter tweaking, would affect the model's performance was the main objective.

SVM Model Summary:

svm_model_summary

Model	Train-test split	Cross validation
Original Features	85.23	83.75
With Hyperparameter Tuning	87.56	85.9
With Feature Selection	86.78	84.85
With PCA	88.12	86.45

This table shows that applying PCA provided the best performance in terms of train-test split accuracy and cross-validation accuracy.

Summary Table of Studio 3: Activity 7

I added more classifiers to our model evaluation in Studio 3, Activity 7: the Multi-Layer Perceptron (MLP) Classifier, RandomForest, and Stochastic Gradient Descent (SGD) Classifier. against determine which model was the best classifier for our task, we compared these models against the SVM.

Classifier Summary:

classifier_summary

Model	Train-test split	Cross validation
SVM	88.12	86.45
SGDClassifier	88.93665806821440	86.64572304696000
RandomForest	91.97477787331610	92.63911807406510
MLPClassifier	77.90197764402410	87.10135222131620

From this summary, the Random Forest model outperformed all others, making it the best overall classifier in this activity.

Systematic approach to develop ML model (Portfolio Task)

Step 1: Data Collection

Objective:

Gathering and preparing the dataset for analysis was the first stage. I concentrated on acceleration data obtained from 17 body-worn sensors that measure accelerations in the x, y, and z directions.

Process:

I took particular body parts' data columns and merged them into a single dataset based on the student number, making sure that every data input was marked as "slicing" (1) or "boning" (0).

- Link to Source Code and Data: Week 3

Step 2: Create Composite Columns

Objective:

I computed new composite features to improve the model's capacity to discriminate between slicing and boning processes.

Process:

I determined the Roll and Pitch angles by computing the Root Mean Square (RMS) values for several axis combinations (x, y, z). After that, the original dataset and these features were combined to create an enriched dataset with 20 columns.

Source Code and Data: Week 3

Step 3: Data Pre-processing and Feature Computation

Objective:

In order to prepare the dataset for machine learning, this stage computed statistical characteristics for every column.

Process:

I computed the following statistical features across a rolling window of sixty frames for each chosen column:

- Mean

- Standard Deviation

- Minimum and Maximum Values

- Area Under the Curve (AUC)

- Number of Peaks

These features were combined to create a comprehensive dataset with 108 features, ready for training the machine learning models.

Link to Source Code and Data: Week 3

Step 4: Model Training

Objective:

use a variety of methods such as feature selection, PCA, hyperparameter tuning, train-test split, cross-validation, and hyperparameter tweaking to train and assess several machine learning models on the given dataset.

Process:

Train-Test Split (70/30):

A 70/30 train-test split was first used to train the SVM model. After that, the model was assessed using the test data, and it received a perfect score.

10-fold Cross-Validation:

Tenfold cross-validation was used to make sure the model's performance remained consistent across data splits. The model's cross-validation score of 0.999907 indicated that it was still operating almost flawlessly.

Hyperparameter Tuning:

We adjusted the SVM model's hyperparameters using RandomizedSearchCV. The model retained the same cross-validation performance and scored flawlessly on the test set even with the smaller parameter grid.

Feature Selection:

We used SelectKBest to narrow the list of features down to the top 10. This smaller feature set was then used to train the adjusted SVM model. There was no discernible decline in precision, and the model continued to operate very flawlessly.

Principal Component Analysis (PCA):

Ultimately, the feature space was reduced to ten major components using PCA. These components were used to train and assess the adjusted SVM model, which continued to produce results that were almost flawless.

Outcome Summary:

	Model	Train-test split	Cross-validation
0	Original Features	1.0	0.999907
1	With Hyperparameter Tuning	1.0	0.999907
2	With Feature Selection	1.0	0.999907
3	With PCA	1.0	0.999907

These findings imply that the SVM model is quite successful, attaining almost flawless accuracy in a variety of setups and validation techniques.

Step 5: Model Selection

Objective:

To compare the SVM configurations against other classifiers, such as SGD, RandomForest, and MLP, and determine which model performs the best.

Process:

We assessed the RandomForest, SVM, SGD, and MLP models using cross-validation as well as a train-test split. Every model performed incredibly well; the majority of them received flawless marks.

Outcome Summary:

	Model	Train-test split	Cross-validation
0	SVM	1.000000	0.999907
1	SGDClassifier	1.000000	0.999815
2	RandomForest	1.000000	1.000000
3	MLPClassifier	0.999383	0.999722

Model Selection:

Best SVM Model:

Since the performance of all SVM configurations was almost the same, other considerations such as processing economy will determine which model is optimal. The SVM with Original Features is advised if interpretability and simplicity are of the utmost importance, as it outperformed the others without necessitating extra preprocessing stages like feature selection or PCA.

Best Overall Model:

The RandomForest model outperformed the others, scoring flawlessly in both the train-test split and cross-validation. It is the greatest option overall for this activity due to its resilience and capacity to manage different levels of data complexity.

Conclusion

For the purpose of this portfolio work, machine learning models for categorizing boning and slicing operations were developed and evaluated using a methodical process. I determined that the RandomForest model was the best performer in this scenario by carefully gathering data, creating features, and training the model. This illustrates the value of ensemble approaches.

I made sure that the results were accurate and dependable by following best practices in model construction and evaluation. This laid a solid platform for my future work in activity recognition.