

CS 6375

ASSIGNMENT _____Lab-2_____

Names of students in your group:

Adrita Dutta

axd172930

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

LAB REPORT

Parameter Tuning of Models on the Wine Dataset

OUTPUT SUMMARY

The following output was achieved at the end of the assignment:

Algorithm	Best Parameters	Average precision	Average recall	Average f1-score	Best score	Accuracy Score
Decision Tree	{'max_features': 'log2', 'min_impurity_decrease': 0.0, 'max_depth': 20, 'min_samples_leaf': 1}	0.95	0.94	0.94	0.922535211268	0.944444 44444444 44
MLP Classifier	{'alpha': 0.0001, 'activation': 'logistic', 'max_iter': 200, 'hidden_layer_sizes': (200, 50)}	0.15	0.39	0.22	0.964788732394	0.388888 88888888 89
SVM Classifier	{'kernel': 'linear', 'C': 1, 'max_iter': -1, 'random_state': 100, 'degree': 3}	0.97	0.97	0.97	0.943661971831	0.972222 22222222 22
Gaussian Naïve Bayes	{'priors': (0.3, 0.4, 0.3)}	0.94	0.94	0.94	0.943661971831	0.944444 44444444 44
Logistic Regression	{'penalty': 'l2', 'C': 1.0, 'max_iter': 100, 'fit_intercept': 'True'}	0.89	0.89	0.89	0.929577464789	0.888888 88888888 88
K-Means Algorithm	{'n_neighbors': 2, 'weights': 'distance', 'algorithm': 'auto', 'p': 1}	0.87	0.86	0.86	0.922535211268	0.861111 11111111 12
Bagging	{'max_features': 3, 'max_samples': 4, 'random_state': 100, 'n_estimators': 80}	0.95	0.94	0.94	0.943661971831	0.944444 44444444 44
Random Forest	{'max_features': 4, 'n_estimators': 80, 'criterion': 'gini', 'max_depth': 3}	0.97	0.97	0.97	0.985915492958	0.972222 22222222 22
AdaBoost	{'n_estimators': 80, 'learning_rate': 0.2, 'random_state': 100, 'algorithm': 'SAMME.R'}	0.97	0.97	0.97	0.929577464789	0.972222 22222222 22
Gradient Boosting	{'max_features': 3, 'n_estimators': 70, 'learning_rate': 0.2, 'max_depth': 3}	1	1	1	0.985915492958	1.0

XGBoost	{'n_estimators': 60, 'learning_rate': 0.2, 'max_delta_step': 1, 'booster': 'gbtree'}	0.95	0.94	0.94	0.950704225352	0.9444444444444444
---------	--	------	------	------	----------------	--------------------

ANALYSIS OF RESULTS

From the classification report and the stats obtained, we can make the following observations:

On the basis of Accuracy Score:

- Gradient Descent Classifier has the best performance with a 100% accuracy and perfect precision and recall values for many classes
- SVM , Random Forest, AdaBoost follows closely behind with an accuracy of 97.2%.
- tuning the neural network, Gradient Boosting overall took comparatively more time as compared to the others
- MLP has the lowest accuracy of 38.8%

On the basis of Best Score:

- Gradient Descent, Random Forst Classifier has the best performance with a 98.5% accuracy and perfect precision and recall values for many classes
- MLP follows closely behind with an accuracy of 96%.
- tuning the neural network, Gradient Boosting overall took comparatively more time as compared to the others
- Decision Tree Classifier has the lowest accuracy of 92.225%

Some steps that could improve the performance:

- Feature selection to pick only the most relevant features
- Use of ensemble methods

Detailed Version of Data Obtained:

1.Decision Tree:

Best score: 0.922535211268

Best parameters set found on development set:

```
{'max_features': 'log2', 'min_impurity_decrease': 0.0, 'max_depth': 20, 'min_samples_leaf': 1}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

```

precision  recall  f1-score  support
1.0      1.00    0.93    0.96    14

```

2.0	0.88	1.00	0.93	14
3.0	1.00	0.88	0.93	8
micro avg	0.94	0.94	0.94	36
macro avg	0.96	0.93	0.94	36
weighted avg	0.95	0.94	0.94	36

Accuracy Score:

0.9444444444444444

Detailed confusion matrix:

```
[[13  1  0]
 [ 0 14  0]
 [ 0  1  7]]
```

2.MLP Classifier:

Best score: 0.964788732394

Best parameters set found on development set:

```
{'alpha': 0.0001, 'activation': 'logistic', 'max_iter': 200, 'hidden_layer_sizes': (200, 50)}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.00	0.00	0.00	14
2.0	0.39	1.00	0.56	14
3.0	0.00	0.00	0.00	8
micro avg	0.39	0.39	0.39	36
macro avg	0.13	0.33	0.19	36
weighted avg	0.15	0.39	0.22	36

Accuracy Score:

0.3888888888888889

Detailed confusion matrix:

```
[[ 0 14  0]
 [ 0 14  0]
 [ 0  8  0]]
```

3.SVM Classifier:

Best score: 0.943661971831

Best parameters set found on development set:

```
{'kernel': 'linear', 'C': 1, 'max_iter': -1, 'random_state': 100, 'degree': 3}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.93	1.00	0.97	14
2.0	1.00	0.93	0.96	14
3.0	1.00	1.00	1.00	8
micro avg	0.97	0.97	0.97	36
macro avg	0.98	0.98	0.98	36
weighted avg	0.97	0.97	0.97	36

Accuracy Score:

0.9722222222222222

Detailed confusion matrix:

```
[[14 0 0]
 [ 1 13 0]
 [ 0 0 8]]
```

4.Gaussian Naïve Bayes:

Best score: 0.943661971831

Best parameters set found on development set:

```
{'priors': (0.3, 0.4, 0.3)}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.93	0.93	0.93	14
2.0	0.93	0.93	0.93	14
3.0	1.00	1.00	1.00	8
micro avg	0.94	0.94	0.94	36

macro avg	0.95	0.95	0.95	36
weighted avg	0.94	0.94	0.94	36

Accuracy Score:

0.9444444444444444

Detailed confusion matrix:

```
[[13  1  0]
 [ 1 13  0]
 [ 0  0  8]]
```

5.Logistic Regression:

Best score: 0.929577464789

Best parameters set found on development set:

```
{'penalty': 'l2', 'C': 1.0, 'max_iter': 100, 'fit_intercept': 'True'}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.86	0.86	0.86	14
2.0	0.86	0.86	0.86	14
3.0	1.00	1.00	1.00	8
micro avg	0.89	0.89	0.89	36
macro avg	0.90	0.90	0.90	36
weighted avg	0.89	0.89	0.89	36

Accuracy Score:

0.8888888888888888

Detailed confusion matrix:

```
[[12  2  0]
 [ 2 12  0]
 [ 0  0  8]]
```

6.K Neighbors Classifier:

Best score: 0.922535211268

Best parameters set found on development set:

```
{'n_neighbors': 2, 'weights': 'distance', 'algorithm': 'auto', 'p': 1 }
```

The model is trained on the full development set.
The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.86	0.86	0.86	14
2.0	0.80	0.86	0.83	14
3.0	1.00	0.88	0.93	8
micro avg	0.86	0.86	0.86	36
macro avg	0.89	0.86	0.87	36
weighted avg	0.87	0.86	0.86	36

Accuracy Score:

0.8611111111111112

Detailed confusion matrix:

```
[[12  2  0]
 [ 2 12  0]
 [ 0  1  7]]
```

7.Bagging:

Best score: 0.943661971831

Best parameters set found on development set:

{'max_features': 3, 'max_samples': 4, 'random_state': 100, 'n_estimators': 80}

The model is trained on the full development set.
The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.88	1.00	0.93	14
2.0	1.00	0.86	0.92	14
3.0	1.00	1.00	1.00	8
micro avg	0.94	0.94	0.94	36
macro avg	0.96	0.95	0.95	36
weighted avg	0.95	0.94	0.94	36

Accuracy Score:

0.9444444444444444

Detailed confusion matrix:

```
[[14 0 0]
 [ 2 12 0]
 [ 0 0 8]]
```

8.RandomForest:

Best score: 0.985915492958

Best parameters set found on development set:

```
{'max_features': 4, 'n_estimators': 80, 'criterion': 'gini', 'max_depth': 3 }
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.93	1.00	0.97	14
2.0	1.00	0.93	0.96	14
3.0	1.00	1.00	1.00	8
micro avg	0.97	0.97	0.97	36
macro avg	0.98	0.98	0.98	36
weighted avg	0.97	0.97	0.97	36

Accuracy Score:

0.9722222222222222

Detailed confusion matrix:

```
[[14 0 0]
 [ 1 13 0]
 [ 0 0 8]]
```

9.AdaBoost:

Best score: 0.929577464789

Best parameters set found on development set:

```
{'n_estimators': 80, 'learning_rate': 0.2, 'random_state': 100, 'algorithm': 'SAMME.R'}
```


The model is trained on the full development set.
The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	1.00	1.00	1.00	14
2.0	0.93	1.00	0.97	14
3.0	1.00	0.88	0.93	8
micro avg	0.97	0.97	0.97	36
macro avg	0.98	0.96	0.97	36
weighted avg	0.97	0.97	0.97	36

Accuracy Score:

0.9722222222222222

Detailed confusion matrix:

```
[[14 0 0]
 [ 0 14 0]
 [ 0 1 7]]
```

10.Gradient Boosting:

Best score: 0.985915492958

Best parameters set found on development set:

```
{'max_features': 3, 'n_estimators': 70, 'learning_rate': 0.2, 'max_depth': 3}
```

The model is trained on the full development set.
The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	1.00	1.00	1.00	14
2.0	1.00	1.00	1.00	14
3.0	1.00	1.00	1.00	8
micro avg	1.00	1.00	1.00	36
macro avg	1.00	1.00	1.00	36
weighted avg	1.00	1.00	1.00	36

Accuracy Score:

1.0

Detailed confusion matrix:

```
[[14 0 0]
 [ 0 14 0]
 [ 0 0 8]]
```

11.XGBoost:

Best score: 0.950704225352

Best parameters set found on development set:

```
{'n_estimators': 60, 'learning_rate': 0.2, 'max_delta_step': 1, 'booster': 'gbtree'}
```

The model is trained on the full development set.

The scores are computed on the full evaluation set.

Detailed classification report:

	precision	recall	f1-score	support
1.0	0.88	1.00	0.93	14
2.0	1.00	0.93	0.96	14
3.0	1.00	0.88	0.93	8
micro avg	0.94	0.94	0.94	36
macro avg	0.96	0.93	0.94	36
weighted avg	0.95	0.94	0.94	36

Accuracy Score:

0.9444444444444444

Detailed confusion matrix:

```
[[14 0 0]
 [ 1 13 0]
 [ 1 0 7]]
```