

CENG499 Assignment3

Uygar Yaşar 2310613

December 2022

1 Part1

```
[ ]
['Outlook']
['Outlook']
['Outlook', 'Windy']
['Outlook', 'Windy']
['Outlook']
['Outlook', 'Humidity']
['Outlook', 'Humidity']
Training completed
Accuracy : 100.00
```

Figure 1: Decision Tree Output With Used Attributes Printing

As it can be seen from the figure 1, Outlook has most information gain and also when I change the criterion I saw that it has also highest gain ratio. Then, Windy and Humidity separates data into leaves. Since all the dataset separate correctly, accuracy is 100.

2 Part2

I used accuracy as performance metric. For dataset1, I set the C values to 1, 5, 10, and 20 and kernel functions were linear, polynomial, and rbf. So, I get 12 plots. (figure 2-5) For dataset2, I printed all the configurations, their accuracies and confidence intervals (It can be seen from figure 6.)

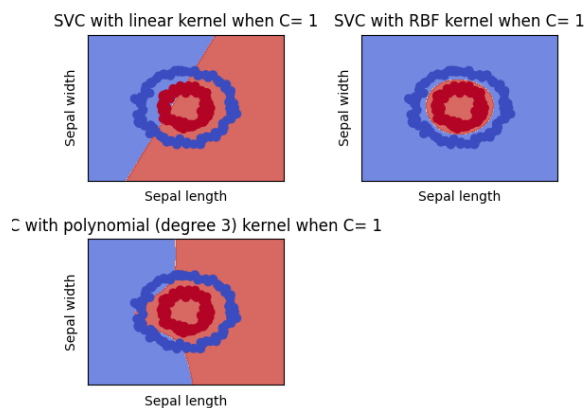


Figure 2: $C=1$ plots

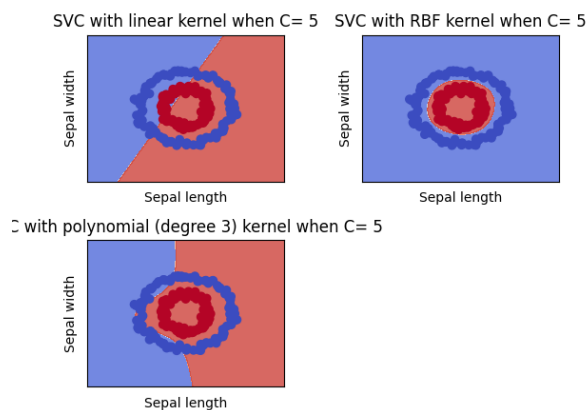


Figure 3: $C=5$ plots

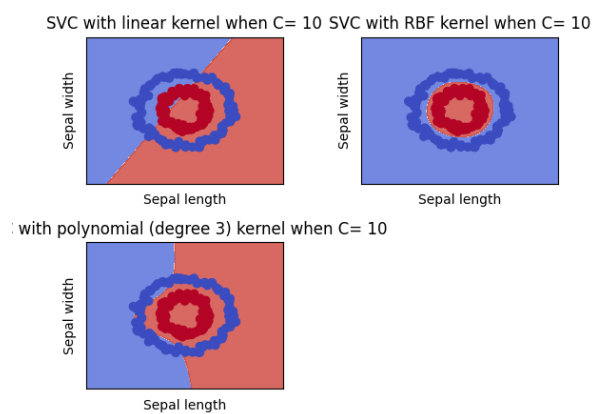


Figure 4: $C=10$ plots

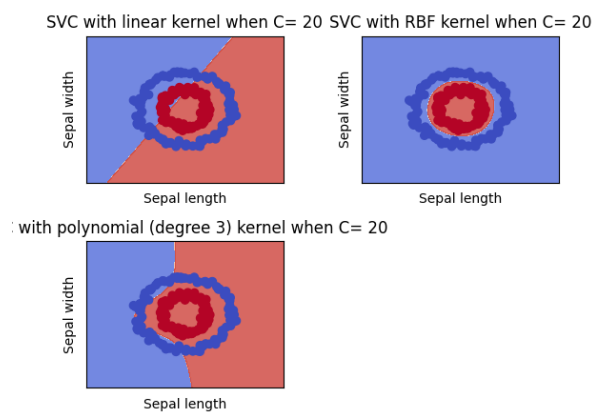


Figure 5: $C=20$ plots

```

C: 0.1 Kernel: rbf Mean: 0.8311999999999999 Confidence interval: [0.8305220029498588, 0.8318779970501411]
C: 0.1 Kernel: poly Mean: 0.7574666666666666 Confidence interval: [0.7552180048423809, 0.7597153284909524]
C: 0.1 Kernel: linear Mean: 0.8752000000000001 Confidence interval: [0.8724880117994358, 0.8779119882005644]
C: 0.1 Kernel: sigmoid Mean: 0.8315999999999999 Confidence interval: [0.8285324407096194, 0.8346675592903804]
C: 1 Kernel: rbf Mean: 0.9225333333333333 Confidence interval: [0.9216031290546933, 0.9234635376119733]
C: 1 Kernel: poly Mean: 0.8762666666666666 Confidence interval: [0.873243502761087, 0.8792898305722463]
C: 1 Kernel: linear Mean: 0.9545333333333336 Confidence interval: [0.95271705141689, 0.9563496152497771]
C: 1 Kernel: sigmoid Mean: 0.8183999999999999 Confidence interval: [0.8151818763230727, 0.8216181236769271]
C: 10 Kernel: rbf Mean: 0.9408000000000001 Confidence interval: [0.9393570585597468, 0.9422429414402533]
C: 10 Kernel: poly Mean: 0.9273333333333333 Confidence interval: [0.9250967287639023, 0.9295699379027644]
C: 10 Kernel: linear Mean: 0.9544 Confidence interval: [0.9535298735724046, 0.9552701264275955]
C: 10 Kernel: sigmoid Mean: 0.7795999999999998 Confidence interval: [0.7767852673306332, 0.7824147326693665]
C: 100 Kernel: rbf Mean: 0.9202666666666669 Confidence interval: [0.9183002689449432, 0.9222330643883906]
C: 100 Kernel: poly Mean: 0.9398666666666667 Confidence interval: [0.9373086049004069, 0.9424247284329266]
C: 100 Kernel: linear Mean: 0.9496000000000002 Confidence interval: [0.9479721425123804, 0.9512278574876201]
C: 100 Kernel: sigmoid Mean: 0.7655999999999998 Confidence interval: [0.7598729134806003, 0.7713270865193393]

```

Figure 6: Dataset 2 result with different configurations

3 Part3

Figure 7 shows the configurations that I used.

```

configs = [
    {"kneighborsclassifier__metric": ["euclidean", "manhattan"],
     "kneighborsclassifier__n_neighbors": [4, 6]},

    {"svc__C": [0.1, 1],
     "svc__kernel": ["poly", "rbf", "linear"]},

    {"decisiontreeclassifier__max_depth": [5, 7],
     "decisiontreeclassifier__criterion": ["gini", "entropy"]},

    {"randomforestclassifier__n_estimators": [7, 15],
     "randomforestclassifier__max_depth": [3, 5]}
]

```

Figure 7: Configurations for all methods

My output shows:

The best parameters for each method.

The accuracy scores and their interval for each.

The accuracy scores and their interval for each as well.

For random forest, I appended best_params_ to a list of dictionaries. The best parameter is probably most occurred element in the list with sized 5.

```

KNN: Accuracy score: 0.7083910257562953 Confidence interval: [ 0.6871435901189943 , 0.7296384613935963 ] F1 score: 0.7961816952822705 Interval of F1: [ 0.779015069383245 , 0.8133483211812959 ]
Best Parameters: {'kneighborsclassifier__metric': 'manhattan', 'kneighborsclassifier__n_neighbors': 6}
SVM: Accuracy score: 0.7517973062883243 Confidence interval: [ 0.7400400709693663 , 0.7635545416072823 ] F1 score: 0.8339288355801167 Interval of F1: [ 0.8221716002611587 , 0.8456860708990747 ]
Best Parameters: {'svc__C': 1, 'svc__kernel': 'rbf'}
Decision Tree: Accuracy score: 0.6981922041802282 Confidence interval: [ 0.6832261314560942 , 0.7131582769043622 ] F1 score: 0.7931007849592476 Interval of F1: [ 0.7789589895720247 , 0.8072425803464706 ]
Best Parameters: {'decisiontreeclassifier__criterion': 'gini', 'decisiontreeclassifier__max_depth': 5}
Random Forest: Accuracy score: 0.7240406873939808 Confidence interval: [ 0.7179910261441979 , 0.7300903486437637 ] F1 score: 0.8294047618215913 Interval of F1: [ 0.8258390579265061 , 0.8329704657166764 ]
Best Parameters: [{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 15}, {'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 7}]

```

Figure 8: Output of Part3

```

Best Parameters: [{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 15},
{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 7},
{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 15},
{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 7},
{'randomforestclassifier__max_depth': 5, 'randomforestclassifier__n_estimators': 7}]

```

Figure 9: The random forest best parameters list (Taken from my output)

As it can be seen from the figure 8 and 9,

Best parameters for KNN: manhattan distance with 6 neighbours.

KNN: Accuracy score: 0.7083910257562953
Confidence interval: [0.6871435901189943 , 0.7296384613935963]
F1 score: 0.7961816952822705
Interval of F1: [0.779015069383245 , 0.8133483211812959]

Best parameters for SVM: rbf with C=1.

SVM: Accuracy score: 0.7517973062883243
Confidence interval: [0.7400400709693663,0.7635545416072823]
F1 score: 0.8339288355801167
Interval of F1: [0.8221716002611587 , 0.8456860708990747]

Best parameters for decision tree: criterion gini and max depth 5.

Decision Tree: Accuracy score: 0.6981922041802282
Confidence interval: [0.6832261314560942 , 0.7131582769043622]
F1 score: 0.7931007849592476
Interval of F1: [0.7789589895720247 , 0.8072425803464706]

For random forest the best parameters are depth 5 and estimators 7 or 15.

Random Forest: Accuracy score: 0.7240406873939808
Confidence interval: [0.7179910261441979 , 0.7300903486437637]
F1 score: 0.8294047618215913
Interval of F1: [0.8258390579265061 , 0.8329704657166764]