CENG499

Assignment 3

Berke Sina Ahlatcı

2468502

Contents

List of Tables

List of Figures

1 Part 1

Here is a console output of the tree created by ID3 algorithm for better visualization.


```
Outlook = overcast
        -> 1

Outlook = rain

-------Windy = false
                -> 1

-------Windy = true
                -> 0

Outlook = sunny

-------Temperature = cool
                -> 1

-------Temperature = hot
                -> 0

-------Temperature = mild

-------------Humidity = high
                    -> 0

-------------Humidity = normal
                    -> 1
```

Figure 1: Console output of the created tree.

2 Part 2

2.1 Dataset 1

I've tried linear and sigmoid as kernel functions, and 1 and 5 as C values.

The following items contain hyperparameter configurations and their corresponding decision boundary plots.
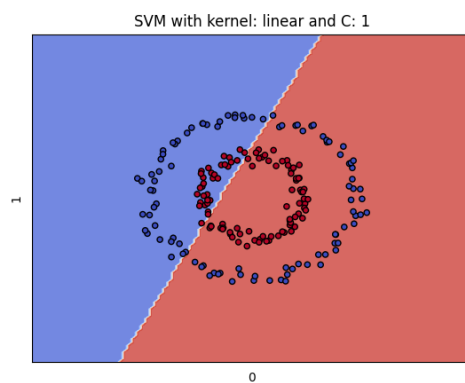
1. Kernel = Linear & C = 1

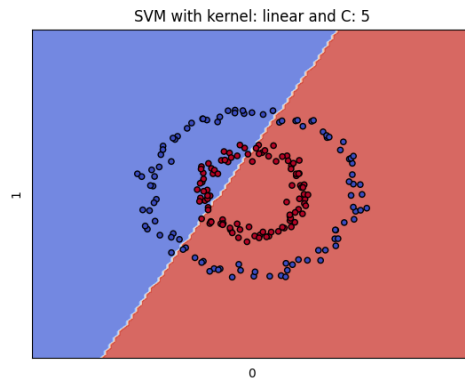Figure 2: Decision boundary of linear & 1.

2. Kernel = Linear & C = 5

SVM with kernel: linear and C: 5

Figure 3: Decision boundary of linear & 5.

3. Kernel = Sigmoid & C = 1

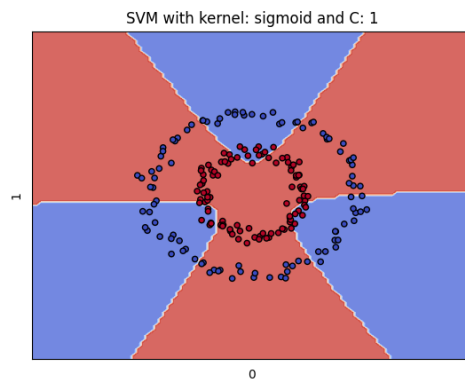SVM with kernel: sigmoid and C: 1

Figure 4: Decision boundary of sigmoid & 1.
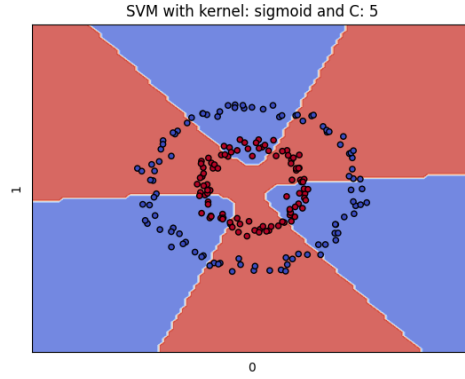
4. Kernel = Sigmoid & C = 5

Figure 5: Decision boundary of sigmoid & 5.

2.2   Dataset 2

I've tried linear and sigmoid as kernel functions, and 1 and 5 as C values. Used the accuracy while determining the best hyperparameter.

Kernel & C	Accuracy in %	Confidence Interval in %
Linear & 1	95.2	0.573
Linear & 5	95.6	0.286
Sigmoid & 1	81.333	0.978
Sigmoid & 5	77.867	1.005

Table 1: Dataset 2 cross-validation procedure results for hyperparameter values.

3 Part 3

I've used the f1 score while choosing the best hyperparameter configuration for each method.

Tried the following configurations for the methods :

- K-nearest neighbors

  - number of neighbors : 1, 5

  - weights : uniform, distance

- Support vector machine

  - kernel : linear, sigmoid

  - C : 1, 5

- Decision Tree

  - criterions : gini, entropy

  - maximum depth : 1, 5

- Random Forest

  - number of estimators : 10, 50

  - criterions : gini, entropy

  - maximum depth : 1, 5

Following table contains F1 scores and confidence intervals of best hyperparameters

Method	Best Parameters	F1 Score	Confidence Interval
KNN	number of neighbors : 5 weights : distance	81.478	0.515
SVM	C : 1 kernel : sigmoid	82.561	0.440
DT	criterion : gini max depth : 1	82.224	0.239
RF	criterion : gini max depth : 5 number of estimators : 50	83.190	0.237

Table 2: Best hyperparameter results for each method.

We can crown random forest method with gini criterion, 5 maximum depth and 50 estimators as the best method among other methods and configurations I've tried for credit application dataset.