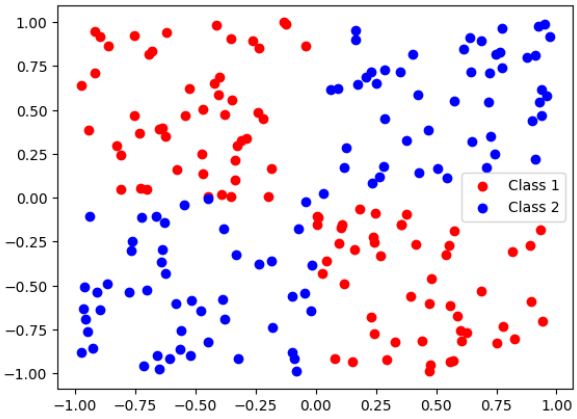
Experiment 4
Decision tree and Random Forest

Name:- Sohini Gupta
Roll No: 20EE38032

March 18, 2024

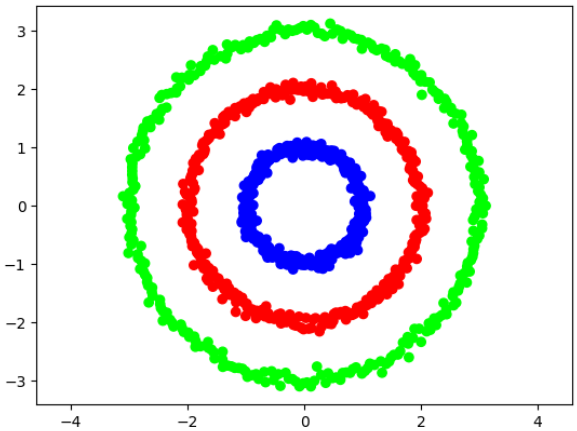
1 Datasets

1.1 Generating two-class dataset (XOR) and visualizing it:

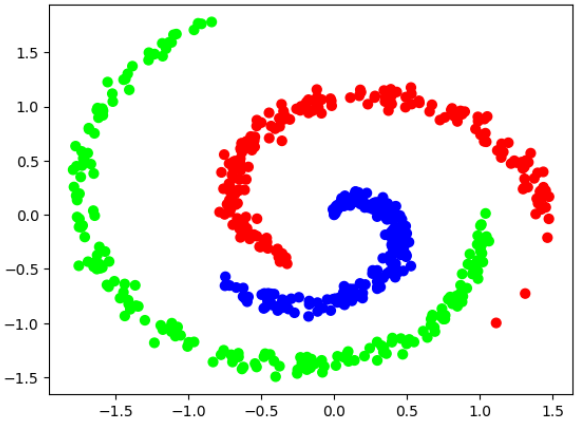


1.2 Generating three-class dataset and visualizing it:

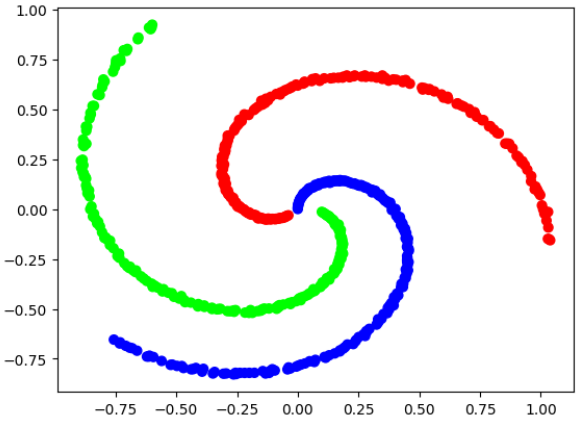
1.2.1 Concentric Circles



1.2.2 Unbalanced Spiral

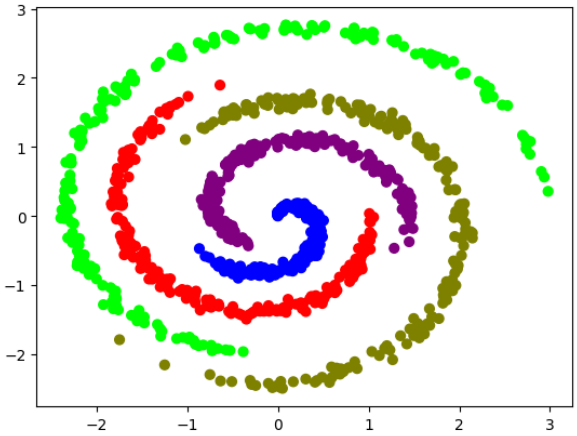


1.2.3 Balanced Spiral

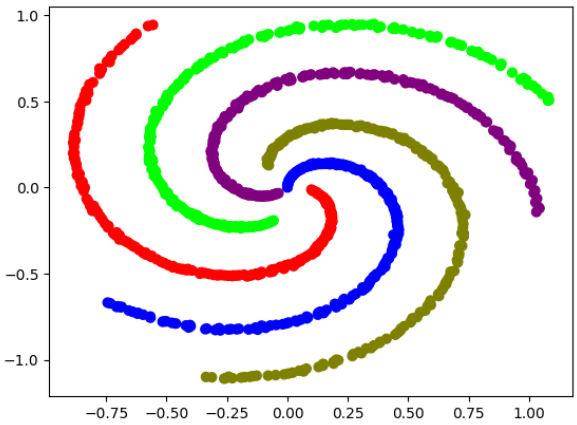


1.3 Generating five class dataset and visualizing it:

1.3.1 Unbalanced Spiral



1.3.2 Balanced Spiral



2 Information Gains, Feature number, threshold values and probability distribution

2.1 For XOR:

```
Decision Tree for XOR Dataset:
[Feature 0 <= 0.94, Info Gain=0.0074]
        [Feature 1 <= 0.37, Info Gain=0.0125]
                [Feature 0 <= -0.06, Info Gain=0.1069]
                        [Feature 1 <= -0.02, Info Gain=0.8344]
                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                        [Feature 1 <= -0.00, Info Gain=0.7045]
                                [Feature 0 <= -0.01, Info Gain=0.1944]
                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= -0.01, Info Gain=0.3054]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                [Feature 0 <= -0.00, Info Gain=0.9580]
                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
        [Feature 1 <= -0.07, Info Gain=0.7793]
                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
```

2.2 For Concentric circles:

```
Decision Tree for Concentric Circles Dataset:
[Feature 0 <= 1.09, Info Gain=0.1634]
        [Feature 0 <= -1.11, Info Gain=0.3259]
                [Feature 0 <= -2.09, Info Gain=0.4766]
                        Leaf: Class=2, Distribution=[0. 0. 1.]
                        [Feature 1 <= -2.23, Info Gain=0.2987]
                                Leaf: Class=2, Distribution=[0. 0. 1.]
                                [Feature 1 <= 1.72, Info Gain=0.6143]
                                        Leaf: Class=1, Distribution=[0. 1. 0.]
                                        Leaf: Class=2, Distribution=[0. 0. 1.]
                [Feature 1 <= -1.61, Info Gain=0.3236]
                        [Feature 1 <= -2.72, Info Gain=0.9722]
                                Leaf: Class=2, Distribution=[0. 0. 1.]
                                Leaf: Class=1, Distribution=[0. 1. 0.]
                        [Feature 1 <= 1.08, Info Gain=0.7828]
                                Leaf: Class=0, Distribution=[1. 0. 0.]
                                [Feature 1 <= 2.09, Info Gain=0.9683]
                                        Leaf: Class=1, Distribution=[0. 1. 0.]
                                        Leaf: Class=2, Distribution=[0. 0. 1.]
        [Feature 0 <= 2.06, Info Gain=0.4468]
                [Feature 1 <= -2.10, Info Gain=0.3060]
                        Leaf: Class=2, Distribution=[0. 0. 1.]
                        [Feature 1 <= 1.78, Info Gain=0.6500]
                                Leaf: Class=1, Distribution=[0. 1. 0.]
                                Leaf: Class=2, Distribution=[0. 0. 1.]
                Leaf: Class=2, Distribution=[0. 0. 1.]
```

2.3 For unbalanced spiral (3 classes)

```
Decision Tree for  Spiral Dataset:
[Feature 1 <= 0.22, Info Gain=0.3106]
        [Feature 1 <= -0.96, Info Gain=0.3386]
                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                [Feature 0 <= -1.33, Info Gain=0.2801]
                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                        [Feature 0 <= 0.52, Info Gain=0.5068]
                                [Feature 0 <= -0.30, Info Gain=0.4298]
                                        [Feature 1 <= -0.46, Info Gain=0.9353]
                                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= 1.07, Info Gain=0.9393]
                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
        [Feature 0 <= -0.87, Info Gain=0.7053]
                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                [Feature 1 <= 1.21, Info Gain=0.1126]
                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
```

2.4 For Balanced spiral (3 classes)

```
Decision Tree for Balanced Spiral Dataset:
[Feature 1 <= 0.15, Info Gain=0.2950]
        [Feature 0 <= 0.19, Info Gain=0.2354]
                [Feature 1 <= -0.06, Info Gain=0.3474]
                        [Feature 1 <= -0.65, Info Gain=0.8748]
                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0.]
                        [Feature 0 <= -0.04, Info Gain=0.8111]
                                [Feature 0 <= -0.83, Info Gain=0.6610]
                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 1 <= -0.01, Info Gain=0.7219]
                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0.]
                [Feature 0 <= 0.46, Info Gain=0.5714]
                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0.]
                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0.]
        [Feature 0 <= -0.60, Info Gain=0.8089]
                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0.]
                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0.]
```

2.5 For Balanced spiral (5 classes)

```
Decision Tree for Unbalanced Spiral Dataset:
[Feature 1 <= 0.37, Info Gain=0.2510]
        [Feature 0 <= -0.06, Info Gain=0.3086]
                [Feature 1 <= -0.24, Info Gain=0.5562]
                        [Feature 1 <= -0.64, Info Gain=0.9978]
                                [Feature 1 <= -1.09, Info Gain=0.8190]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                        [Feature 0 <= -0.74, Info Gain=0.6395]
                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= -0.32, Info Gain=0.4101]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                        [Feature 1 <= -0.19, Info Gain=0.8571]
                                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                                [Feature 0 <= -0.09, Info Gain=0.4652]
                                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                        [Feature 1 <= -0.04, Info Gain=0.8113]
                                                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                [Feature 0 <= 0.46, Info Gain=0.3753]
                        [Feature 1 <= 0.14, Info Gain=0.4711]
                                [Feature 1 <= -0.85, Info Gain=0.4571]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        [Feature 0 <= 0.19, Info Gain=0.3196]
                                                [Feature 1 <= -0.01, Info Gain=0.5155]
                                                        [Feature 1 <= -0.70, Info Gain=0.6975]
                                                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                                [Feature 0 <= -0.04, Info Gain=0.2382]
                                                                        [Feature 1 <= -0.46, Info Gain=0.9544]
                                                                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                                                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                        [Feature 0 <= 0.73, Info Gain=0.8184]
                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
        [Feature 0 <= -0.59, Info Gain=0.5649]
                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                [Feature 1 <= 0.67, Info Gain=0.4884]
                        [Feature 0 <= -0.42, Info Gain=0.4334]
                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                [Feature 0 <= 0.83, Info Gain=0.4690]
                                        Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
```

2.6 For Unbalanced spiral (5 classes)

```
Decision Tree for Unbalanced Spiral Dataset:
[Feature 0 <= -0.87, Info Gain=0.3266]
        [Feature 0 <= -1.85, Info Gain=0.3081]
                [Feature 0 <= -1.91, Info Gain=0.0510]
                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                        [Feature 0 <= -1.91, Info Gain=0.5917]
                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                [Feature 1 <= -1.29, Info Gain=0.2797]
                        [Feature 1 <= -1.88, Info Gain=0.4715]
                                [Feature 0 <= -0.91, Info Gain=0.7219]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                        [Feature 1 <= 1.79, Info Gain=0.4753]
                                [Feature 0 <= -1.72, Info Gain=0.1663]
                                        [Feature 1 <= -1.18, Info Gain=0.5178]
                                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                                [Feature 1 <= 0.67, Info Gain=0.4395]
                                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                        [Feature 0 <= -0.97, Info Gain=0.1145]
                                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                                [Feature 1 <= -1.11, Info Gain=0.4040]
                                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                                        [Feature 0 <= -0.90, Info Gain=0.9710]
                                                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
        [Feature 1 <= 0.21, Info Gain=0.4332]
                [Feature 1 <= -0.95, Info Gain=0.5248]
                        [Feature 1 <= -1.42, Info Gain=0.6551]
                                [Feature 0 <= 0.17, Info Gain=0.2511]
                                        [Feature 1 <= -2.29, Info Gain=0.9819]
                                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= 0.67, Info Gain=0.6500]
                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                        [Feature 0 <= 0.55, Info Gain=0.6238]
                                [Feature 0 <= -0.33, Info Gain=0.4441]
                                        [Feature 1 <= -0.48, Info Gain=0.9082]
                                                Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=0, Distribution=[1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= 1.08, Info Gain=0.9980]
                                        Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                        [Feature 0 <= 1.48, Info Gain=0.9837]
                                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                                Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                [Feature 1 <= 1.20, Info Gain=0.6031]
                        [Feature 0 <= 1.42, Info Gain=0.7549]
                                Leaf: Class=1, Distribution=[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
                                [Feature 0 <= 2.01, Info Gain=0.6292]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                        [Feature 1 <= 1.77, Info Gain=0.6661]
                                [Feature 0 <= 1.46, Info Gain=0.5822]
                                        Leaf: Class=3, Distribution=[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                [Feature 0 <= -0.85, Info Gain=0.0912]
                                        [Feature 0 <= -0.86, Info Gain=1.0000]
                                                Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
                                                Leaf: Class=2, Distribution=[0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]
                                        Leaf: Class=4, Distribution=[0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
```

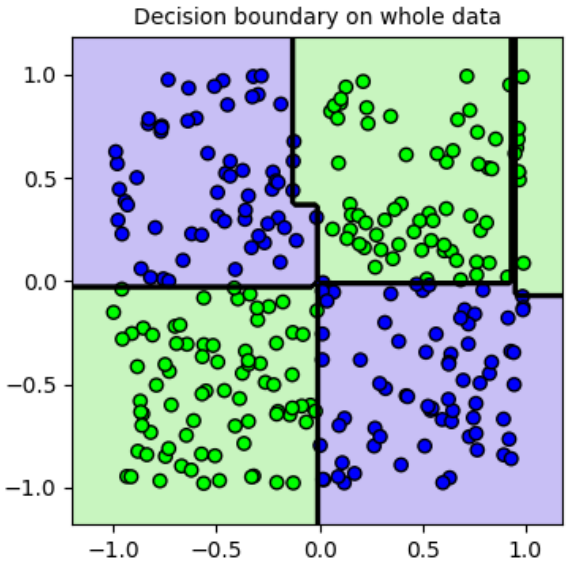
3 Accuracy

- Accuracy for XOR Dataset: 0.97
- Accuracy for Concentric Circles Dataset(3 classes): 0.96
- Accuracy for Unbalanced Spiral Dataset(3 classes): 0.97
- Accuracy for Balanced Spiral Dataset(3 classes): 0.97
- Accuracy for Unbalanced Spiral Dataset (5 classes): 0.97
- Accuracy for Balanced Spiral Dataset (5 classes): 0.96

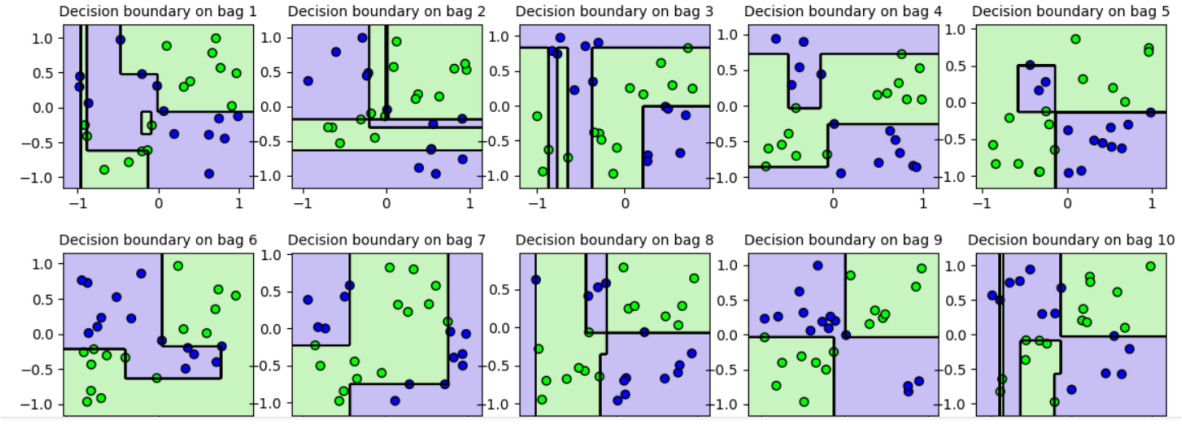
4 Random Forrest

4.1 For XOR

4.1.1 Decision boundaries for one decision tree in 2D space:



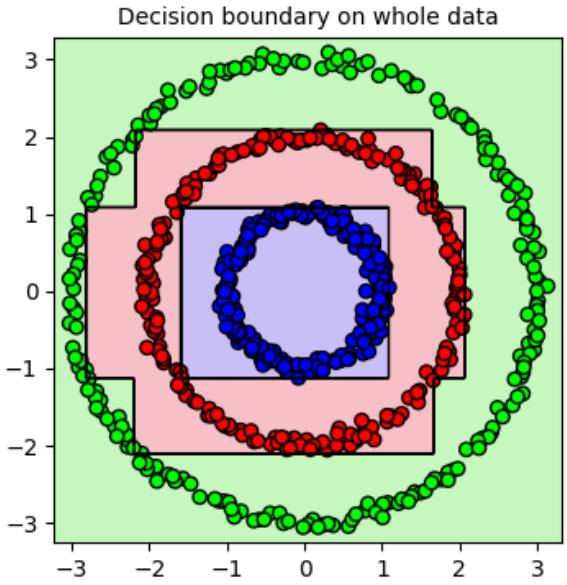
4.1.2 decision boundaries for the random forest with 10 trees on 2D space



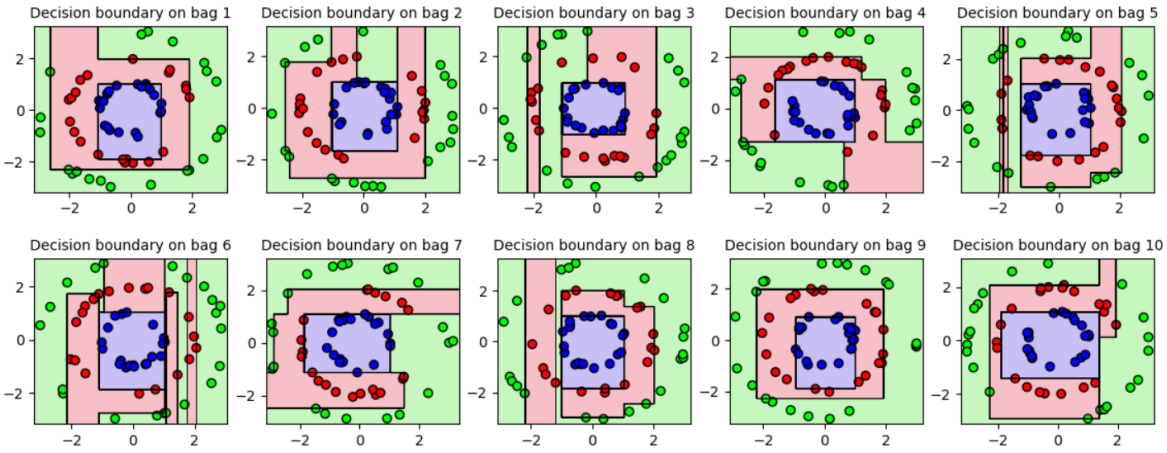
Accuracy in Random forest: 0.8666666666666667

4.2 For concentric circle(3 classes)

4.2.1 Decision boundaries for one decision tree in 2D space:



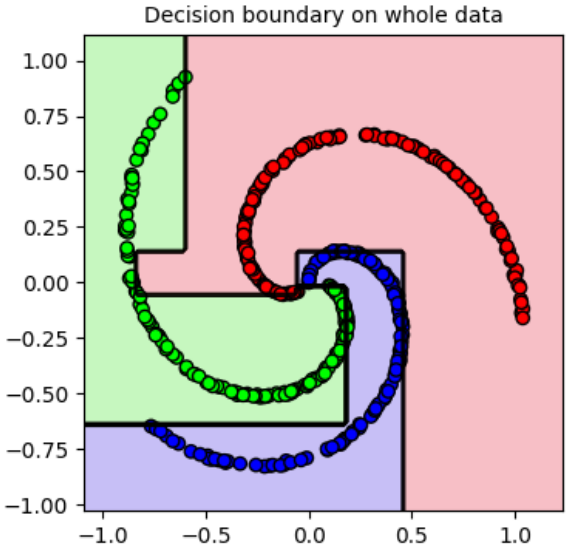
4.2.2 Decision boundaries for the random forest with 10 trees on 2D space



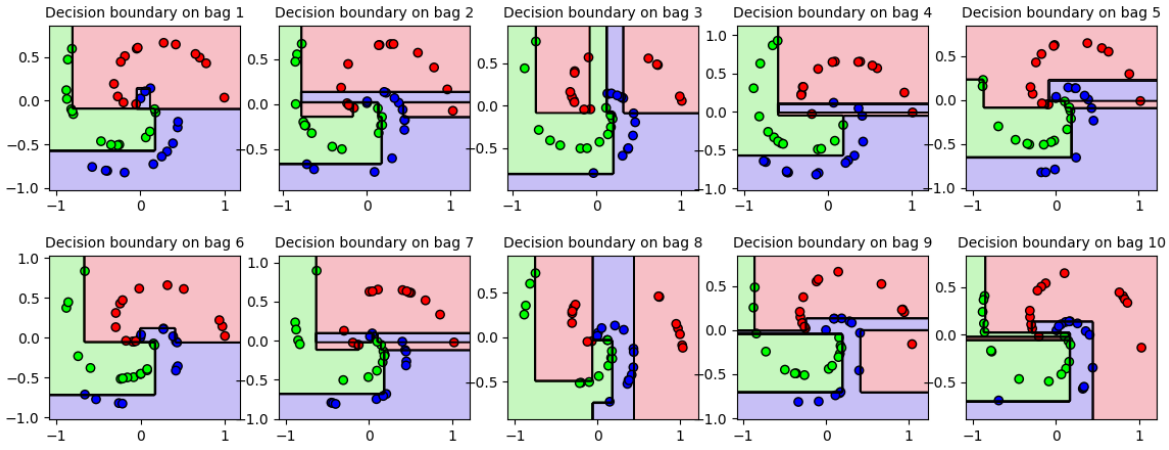
Accuracy in Random forest: 0.9481481481481482

4.3 For balanced spiral (3 classes)

4.3.1 Decision boundaries for one decision tree in 2D space:



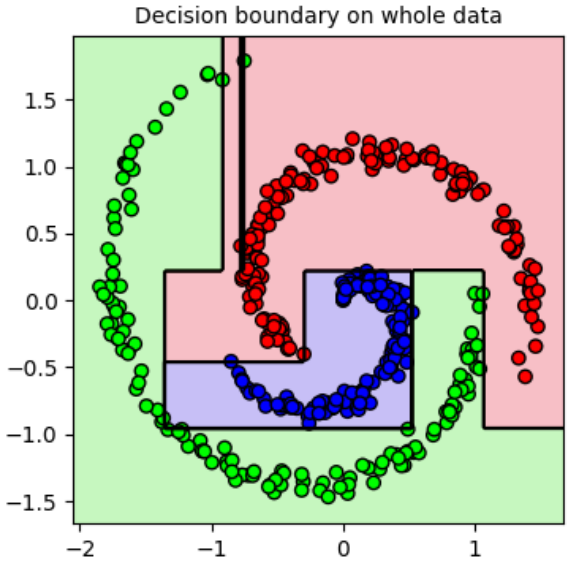
4.3.2 Decision boundaries for the random forest with 10 trees on 2D space



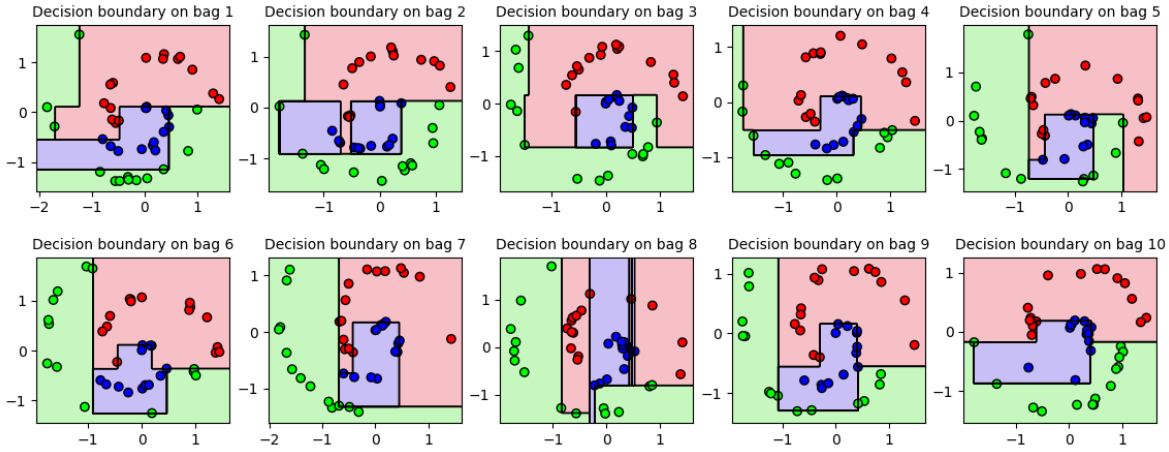
Accuracy in Random forest: 0.8722222222222222

4.4 For unbalanced Spiral(3 classes)

4.4.1 Decision boundaries for one decision tree in 2D space:



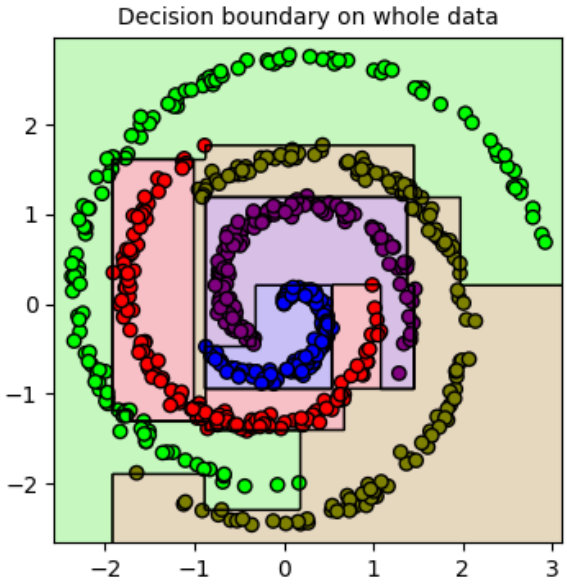
4.4.2 Decision boundaries for the random forest with 10 trees on 2D space



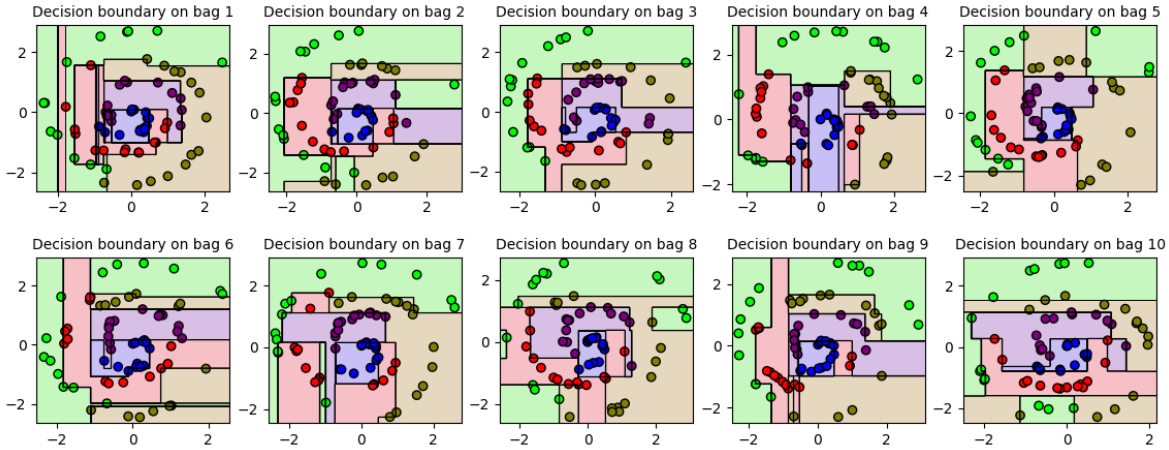
Accuracy in Random forest: 0.8833333333333333

4.5 For unbalanced spiral(5 classes)

4.5.1 Decision boundaries for one decision tree in 2D space:



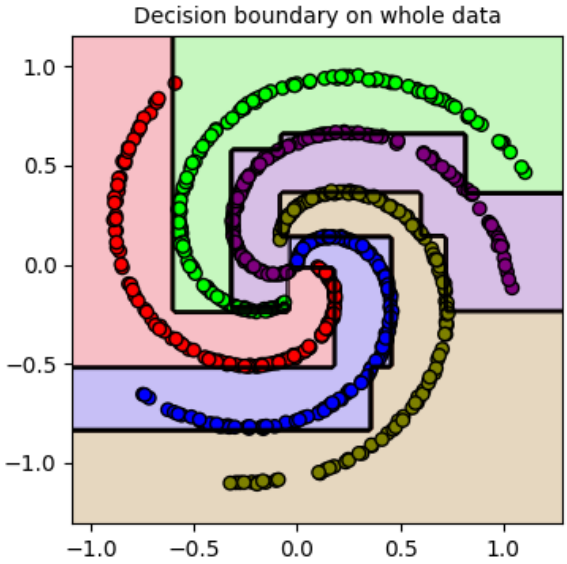
4.5.2 Decision boundaries for the random forest with 10 trees on 2D space



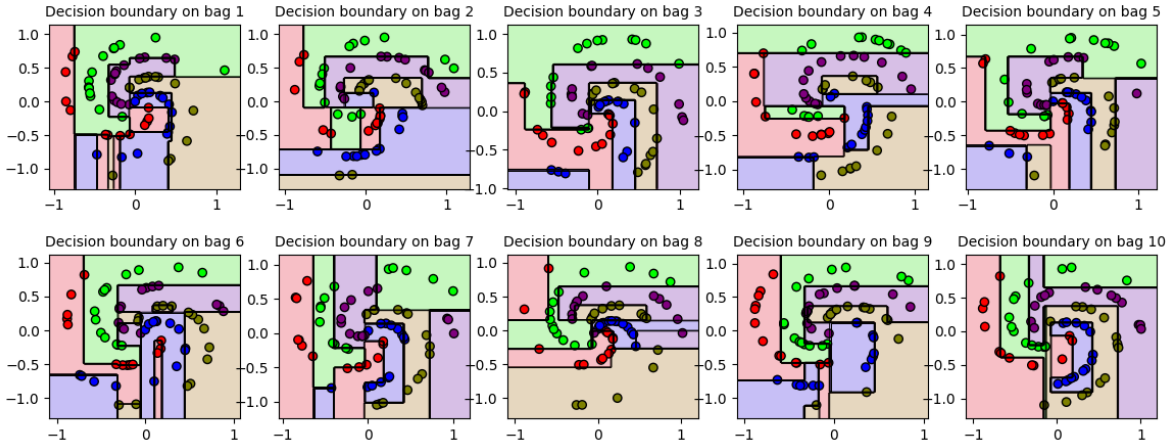
Accuracy in Random forest: 0.82

4.6 For balanced Spiral (5 classes)

4.6.1 Decision boundaries for one decision tree in 2D space:



4.6.2 Decision boundaries for the random forest with 10 trees on 2D space



Accuracy in Random forest: 0.85