Machine Learning Assignment 3

goktugekinci5

December 2022

Part 1

I completed the required functions in the regarding file. My accuracy is outputs 100.0%. My tree is exiting if the given criterion is not either "information gain" or "gain ratio".

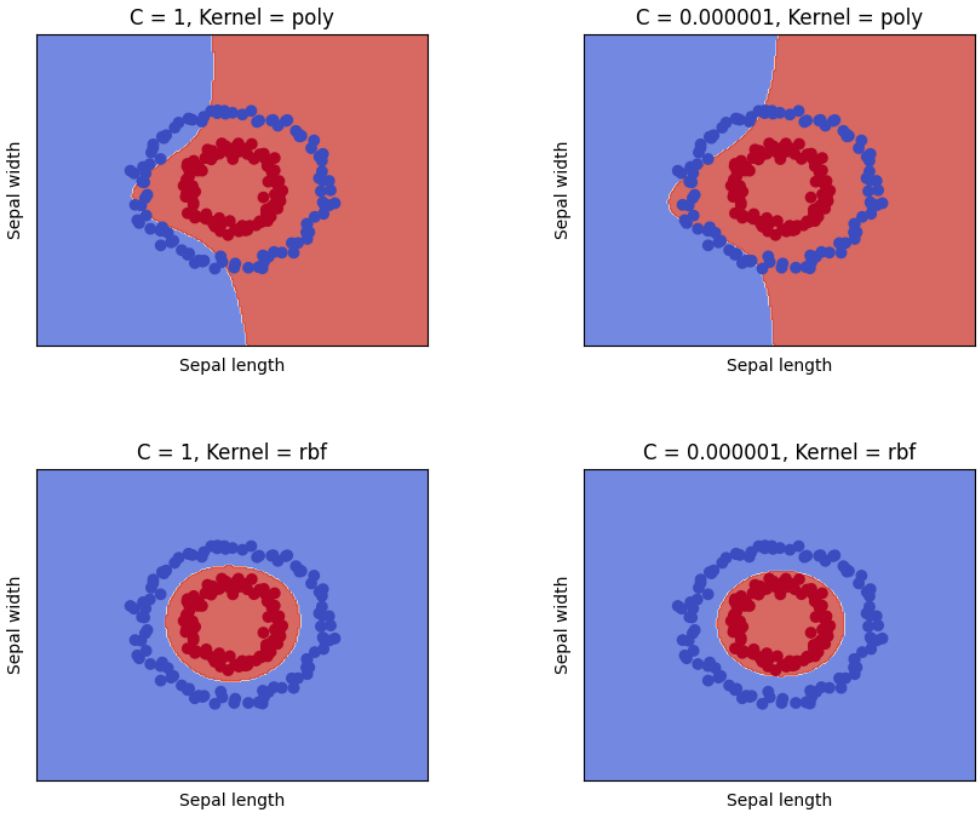
Part 2

Dataset 1

For this dataset, I performed a grid search over the svm models. My configurations is as below:

$$C = \{0.000001,1\}$$
$$\text{Kernel Function} = \{\text{poly, rbf}\}$$

Here are the resulting plots of the configurations:



Dataset 2

In this part I created the configurations for the SVM models regarding the dataset. I tried 4 different configurations for the grid. Here is the results :

$$C = \{0.1,1,10,100\}$$
$$\text{Kernel Function} = \{\text{linear, sigmoid, poly, rbf}\}$$

Here are the results of the configurations.


```
{
    "C": 0.1,
    "kernel": "linear",
    "Confidence Interval": [
        0.8598,
        0.8835
    ]
}
{
    "C": 0.1,
    "kernel": "sigmoid",
    "Confidence Interval": [
        0.8253,
        0.8413
    ]
}
{
    "C": 0.1,
    "kernel": "rbf",
    "Confidence Interval": [
        0.8179,
        0.8288
    ]
}
```

```
{
    "C": 0.1,
    "kernel": "poly",
    "Confidence Interval": [
        0.74,
        0.77
    ]
}
{
    "C": 1,
    "kernel": "linear",
    "Confidence Interval": [
        0.9503,
        0.9663
    ]
}
{
    "C": 1,
    "kernel": "sigmoid",
    "Confidence Interval": [
        0.7853,
        0.848
    ]
}
```

```
{
    "C": 1,
    "kernel": "rbf",
    "Confidence Interval": [
        0.9034,
        0.9232
    ]
}
{
    "C": 1,
    "kernel": "poly",
    "Confidence Interval": [
        0.8508,
        0.8859
    ]
}
{
    "C": 10,
    "kernel": "linear",
    "Confidence Interval": [
        0.9512,
        0.9621
    ]
}
```

```
{
    "C": 10,
    "kernel": "sigmoid",
    "Confidence Interval": [
        0.766,
        0.8006
    ]
}
{
    "C": 10,
    "kernel": "rbf",
    "Confidence Interval": [
        0.9274,
        0.9493
    ]
}
{
    "C": 10,
    "kernel": "poly",
    "Confidence Interval": [
        0.9125,
        0.9408
    ]
}
```

```
{
    "C": 100,
    "kernel": "rbf",
    "Confidence Interval": [
        0.9129,
        0.9238
    ]
}
{
    "C": 100,
    "kernel": "poly",
    "Confidence Interval": [
        0.9216,
        0.9517
    ]
}
```

Part 3

In this part, I tried four different algorithms and various configurations for each of them. Here are the configurations.
For SVM:

$$C = \{0.1, 1, 10\}$$
$$\text{Kernel Function} = \{\text{rbf,linear,poly}\}$$

For KNN:

$$\text{N Neighbors: } \{2,3,4,5,6\},$$
$$\text{Distance Metric: } \{\text{euclidean, manhattan,cosine}\}$$

For Decision Tree:

$$\text{Max Depth: } \{3,6,9\},$$
$$\text{Distance Metric: } \{\text{gini, entropy}\}$$

For Random Forest:

$$\text{N Estimators: } \{10,20,30\},$$
$$\text{Max Depth: } \{3,6,9\}$$

Results


```
    "SVM": {
        "Params": {
            "svc__C": 1,
            "svc__kernel": "rbf"
        },
        "Accuracy": [
            0.7388988281377953,
            0.7684345051955379
        ],
        "F1": [
            0.8282668232080738,
            0.8476626381768589
        ]
    },
    "KNN": {
        "Params": {
            "kneighborsclassifier__metric": "manhattan",
            "kneighborsclassifier__n_neighbors": 5
        },
        "Accuracy": [
            0.7121037981208291,
            0.7398962018791706
        ],
        "F1": [
            0.8095770358812102,
            0.8275658925413525
        ]
    },
```

```
    "Decision Tree": {
        "Params": {
            "decisiontreeclassifier__criterion": "entropy",
            "decisiontreeclassifier__max_depth": 3
        },
        "Accuracy": [
            0.67246754721093,
            0.7021991194557364
        ],
        "F1": [
            0.7772022376159553,
            0.8009193821006533
        ]
    },
    "Random Forest": {
        "Params": {
            "randomforestclassifier__max_depth": 9,
            "randomforestclassifier__n_estimators": 20
        },
        "Accuracy": [
            0.7310498318226385,
            0.7562835015106947
        ],
        "F1": [
            0.8269449376370777,
            0.8430910738884345
        ]
    }
}
```

Here you can see each configuration an their confidence intervals. According to the accuracy scores, SVM is better than all. It is also seen that f1 scores are higher than the accuracies, naturally. For the F1 scores, again SVM is better but random forest is also competetive.