CS170 Intro to AI

Project 2: Feature Selection with Nearest Neighbor

Student Name1: **Abhinav Allam** SID: **862200322** Lecture Session: **002**

Student Name2: **Jacqueline Gardea** SID: **862305608** Lecture Session: **001**

Student Name3: **Majd Kawak** SID: **862273310** Lecture Session: **001**

Student Name4: **Rovin Soriano** SID: **862149089** Lecture Session: **001**

Solution:

Dataset	Best Feature Set	Accuracy
Small Number: <CS170_Spring_2023_Small_data__44.txt>	Forward Selection (Normalized= {3, 5}	0.95
	Backward Elimination = {1, 2, 3, 7}	0.90
	Custom Algorithm = Random Feature Combination (one run, will differ each time) = {1, 3, 5} 10 subsets with 3 features each set Features sets: [(1, 3, 5), (3, 4, 9), (2, 9, 10), (1, 2, 5), (1, 5, 8), (1, 6, 8), (4, 5, 8), (4, 5, 7), (3, 4, 8), (4, 6, 10)]	0.86
Large Number: <CS170_Spring_2023_Large_data__44.txt>	Forward Selection = {14, 10}	0.96
	Backward Elimination = {} Default Rate	0.85
	Custom Algorithm = Random Feature Combination (one run*, will differ each time) = {7, 10, 14, 15} 40 subsets with 4 features each set Features sets: To large to fit	0.91

----------<Begin Report>----------

In completing this project, we consulted following resources:

In completing this project, we utilized a variety of resources to enhance our understanding and develop effective solutions. Firstly, we relied on the lecture slides and materials provided during the course. These resources served as a foundation for our learning and provided valuable insights into the concepts and techniques related to feature selection algorithms. Additionally, we leveraged online sources such as YouTube tutorials and other classes(172, 173) to gain alternative perspectives and deepen our knowledge. By combining these resources, we were able to approach the project with a well-rounded understanding and implement effective strategies in my code.

Contribution of each student in the group:

Jacqueline Gardea

- Implemented the Forward Selection Algorithm
- Analyzing code and running time(Time complexity)
- Tested and run given files

Majd Kawak

- Implemented the Forward Selection Algorithm
- Implemented the Special Algorithm (Random Feature Combinations)
- Implemented the LOOCV function
- Tested and run given files

Rovin Soriano

- Implemented the Backward Elimination Algorithm
- Analyzing code and running time(Time complexity)
- Tested and run given files

Abhinav Allam

- Implemented the Backward Elimination Algorithm
- Implemented the LOOCV function
- Tested and run given files

I. Introduction

This code implements feature selection algorithms developed by Abhinav Allam, Jacqueline Gardea, Majd Kawak, and Rovin Soriano. The algorithms include Forward Selection, Backward Elimination, and Random Combinations. The code takes a dataset as input and performs LOOCV to evaluate the accuracy of different feature subsets. The goal is to identify the optimal feature set that leads to the highest classification accuracy. Additionally, the code provides options for normalizing the data using Z normalization. The selected algorithm can be chosen by the user. Overall, these feature selection algorithms enable efficient and effective feature subset evaluation for classification tasks.

II. Challenges

Some challenges we encountered:

1. Complexity: One of the main challenges we encountered was dealing with the complexity of the algorithms. Understanding the underlying concepts and implementing them correctly required a solid understanding of writing efficient code(Time complexity) and data analysis.
2. Data Preprocessing: We found that data preprocessing was a significant challenge. Handling .txt files, dealing with categorical features, and normalizing the data was a somewhat complex process.
3. Overfitting : It was important to select features that generalized well to new data without sacrificing important information. Finding the right balance required careful model selection and validation, especially for our special algorithm.
4. Programming in Python presented some challenges for us as we were new to the language. Understanding the syntax and properly structuring our code initially posed difficulties, but we were able to overcome these obstacles and gradually improve our proficiency in Python programming.

III. Code Design

Our code demonstrates a well-structured design that promotes code reusability and readability. It follows good coding practices such as object oriented programming, modularization, encapsulation, and separation of concerns. Our code is organized into several functions, each responsible for a specific task, such as leave-one-out cross-validation, feature selection algorithms (forward selection, backward elimination, and random combinations), and data normalization. This modular design allows for easy understanding and maintenance of the code. Additionally, meaningful variable names and comments throughout the code further enhance its readability and make it easier for others to comprehend and modify the code.

IV. Dataset details

The General Small Dataset: Number of features, number of instances:

Number of Features: 10

Number of Instances: 100

The General Large Dataset: Number of features, number of instances:

Number of Features: 40

Number of Instances: 1000

Your Personal Small Dataset: Number of features, number of instances:

Number of Features: 10

Number of Instances: 100

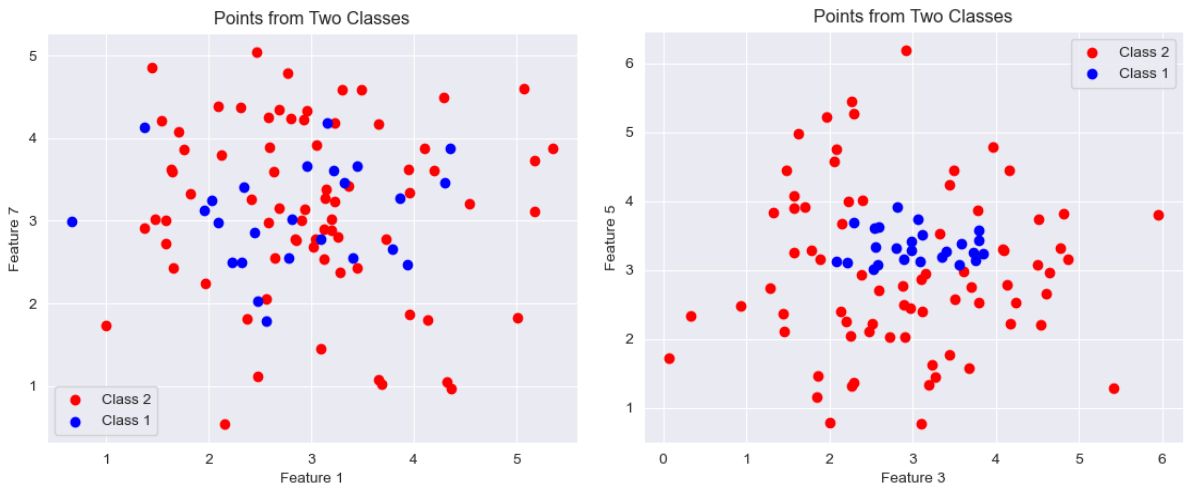
Your Personal Large Dataset: Number of features, number of instances:

Number of Features: 40

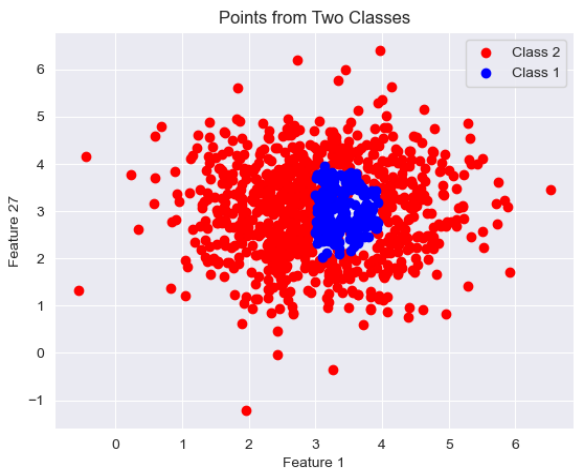
Number of Instances: 1000

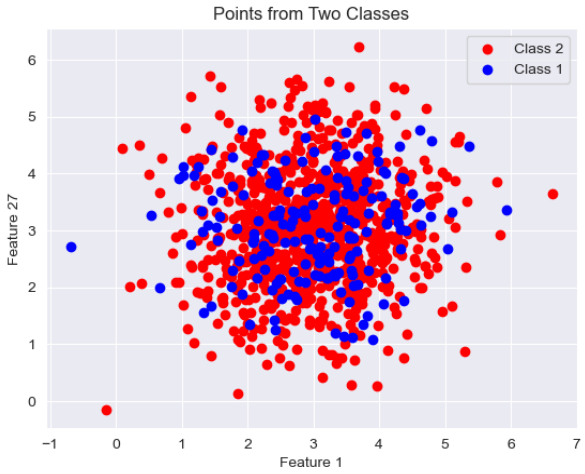
Plot some features and color code them by class and explore your dataset.

General Small Dataset:

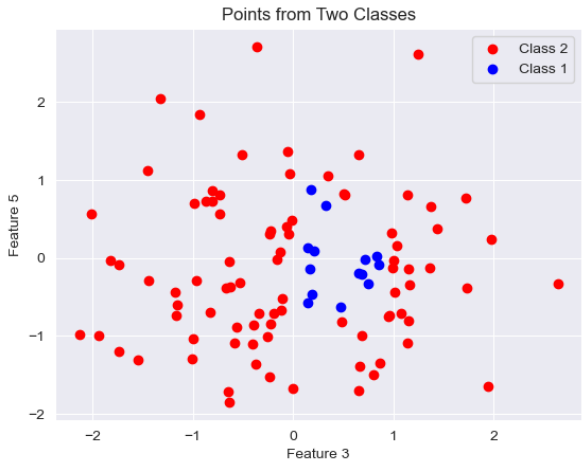


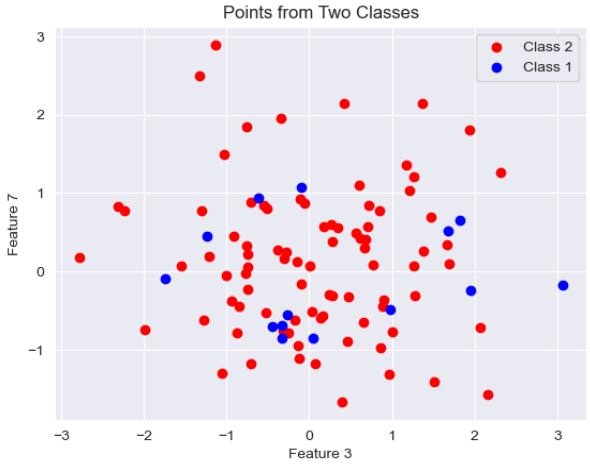
General Large Dataset:



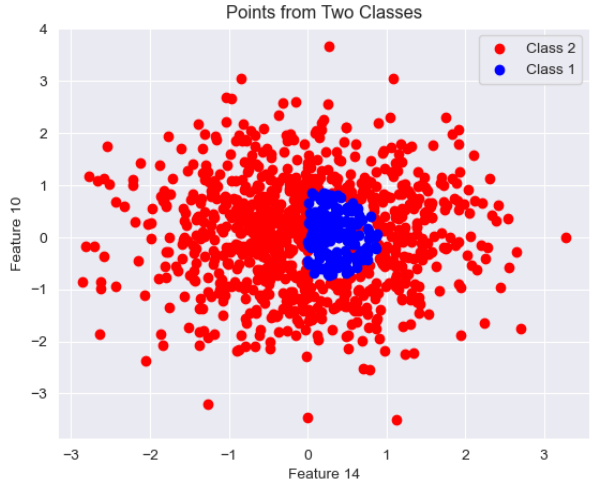


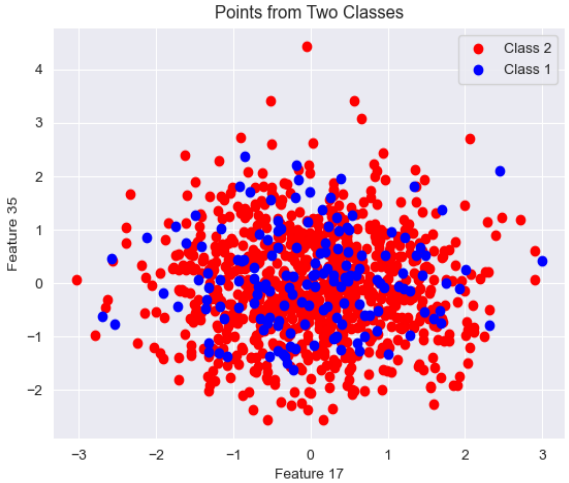
Personal(44) Small Dataset:





Personal(44) Large Dataset:





V. Algorithms

1. Forward Selection

The Forward Selection Algorithm starts by calculating the default rate of the data set with no features selected. It starts the forward feature selection, with no features selected, by iterating through each feature, and determining its accuracy by using leave_one_out_cross_validation. The feature that is determined to have the highest accuracy is added to the feature set. It repeats the process of the forward this time with a feature set that had the highest accuracy from the previous iteration. The algorithm finishes once all features have been added to the feature set. It will display a warning if the accuracy has decreased and at the end display the feature set with the highest accuracy.

2. Backward Elimination

The Backward Selection Algorithm starts with all the features in the model and then we iteratively remove one feature at a time. The feature being eliminated is the one that least decreases the performance of the model. This keeps on iterating with the remaining features until we reach the point where eliminating additional features no longer improves or maintains the performance of the model.

3. Random Feature Combinations

The Random Feature Combinations Algorithm implements a special algorithm, where it generates a specified number of random feature sets, each containing a specified number of features. It then performs Leave One Out Cross Validation on each feature set and calculates the accuracy. The algorithm selects the feature set with the highest accuracy as the optimal set.

VI. Analysis

Experiment 1: Comparing Forward Selection vs Backward Elimination.

The accuracy for the Small dataset was 95% using forward selection and 90% using backward elimination. For the Large dataset, the accuracy was 96% using forward selection and 85% using backward elimination.

Forward selection has shown improved feature selection compared to backward elimination in many cases. While backward elimination may result in higher default rate accuracy output due to its conservative nature, forward selection generally offers better performance and can lead to more accurate models in various scenarios. Therefore, incorporating forward selection as a feature selection technique in our program enhanced the model's predictive power and improved its ability to identify relevant features. By iteratively adding features based on their individual contributions to the model's performance and running time, forward selection allows for a more thorough exploration of the feature space and enables the identification of important predictors. This iterative approach often resulted in a more optimal subset of features, leading to improved accuracy and generalization capabilities of the model.

Experiment 2: Effect of normalization, Compare accuracy when using normalized vs unnormalized data.

We investigated the effect of normalization on accuracy by comparing results obtained from using normalized data versus unnormalized data. Normalization, although it is known to be beneficial for many machine learning algorithms, it introduced additional processing time by a large scale for large datasets. Additionally, depending on the dataset, normalization led to different best subsets of selected features. Therefore, by analyzing the accuracy obtained with and without normalization, we were able to understand the impact of this preprocessing step on the performance of the feature selection algorithms.

VII. Conclusion

General summary of your findings from the Analysis. Potential improvements to this approach of doing feature selection.

Analysis:

Overall, our experiments highlighted the importance of selecting appropriate feature selection techniques and considering the impact of preprocessing steps, such as normalization, on the performance of the algorithms. Forward Selection has proven to be a more effective method for feature selection compared to Backward Elimination, consistently yielding higher accuracy. Additionally, the choice of normalization can significantly influence the results and should be carefully considered based on the specific dataset and requirements of the problem at hand.

Potential Improvements:

1. **Implement Different Classifiers**: Our current implementation focuses on utilizing the k-nearest neighbor (k-NN) algorithm for feature selection. However, to offer a wider range of options for users, it would be beneficial to incorporate the ability to choose from various classifiers, such as decision trees, or logistic regression. This would enable us to explore different classification algorithms and assess their effectiveness in feature selection based on their specific data and requirements.
2. **Include Different Validation Methods**: Although our program employs the LOOCV method, the integration of additional validation techniques could be beneficial. For example, incorporating Holdout Validation where we split data into 2 sets(training and testing), K-Fold Cross-Validation we learned in class, or Repeated Cross-Validation. Those would enable a more comprehensive evaluation of the selected features' stability and generalization capability across the datasets.

VIII. Trace of your small dataset

Small-test-dataset.txt forward selection:

```
Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection
Algorithms.
Type in the name of the file to test: small-test-dataset.txt
File processed successfully.
Type the number of the algorithm you want to run.
1- Forward Selection.
2- Backward Elimination.
3- Random Combinations.
1

This dataset has 10 features (not including the class attribute), with 100
instances.
Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Running nearest neighbor with no features (default rate), using
“leaving-one-out” evaluation, we get an accuracy of  75.0 %
--------------------------------------------
Beginning search (Forward Feature Selection).
--------------------------------------------
On level  1  of the search tree

--Considering adding feature number  1  gets us a total accuracy of
56.99999999999999 %
--Considering adding feature number  2  gets us a total accuracy of  54.0 %
--Considering adding feature number  3  gets us a total accuracy of  68.0 %
--Considering adding feature number  4  gets us a total accuracy of  65.0 %
--Considering adding feature number  5  gets us a total accuracy of  75.0 %
--Considering adding feature number  6  gets us a total accuracy of  61.0 %
--Considering adding feature number  7  gets us a total accuracy of  62.0 %
--Considering adding feature number  8  gets us a total accuracy of  60.0 %
--Considering adding feature number  9  gets us a total accuracy of  66.0 %
--Considering adding feature number  10  gets us a total accuracy of  64.0 %

On level  1  feature { 5 } was added to current set
Current Set of Features:  [5]  with accuracy of  75.0 %
--------------------------------------------
On level  2  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  76.0 %
```

```
--Considering adding feature number  2  gets us a total accuracy of  80.0 %
--Considering adding feature number  3  gets us a total accuracy of  92.0 %
--Considering adding feature number  4  gets us a total accuracy of  75.0 %
--Considering adding feature number  6  gets us a total accuracy of  79.0 %
--Considering adding feature number  7  gets us a total accuracy of  80.0 %
--Considering adding feature number  8  gets us a total accuracy of  77.0 %
--Considering adding feature number  9  gets us a total accuracy of  73.0 %
--Considering adding feature number  10  gets us a total accuracy of  83.0 %

On level  2  feature { 3 } was added to current set
Current Set of Features:  [5, 3]  with accuracy of  92.0 %
--------------------------------------------
On level  3  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  84.0 %
--Considering adding feature number  2  gets us a total accuracy of  79.0 %
--Considering adding feature number  4  gets us a total accuracy of  84.0 %
--Considering adding feature number  6  gets us a total accuracy of  82.0 %
--Considering adding feature number  7  gets us a total accuracy of  89.0 %
--Considering adding feature number  8  gets us a total accuracy of  79.0 %
--Considering adding feature number  9  gets us a total accuracy of  83.0 %
--Considering adding feature number  10  gets us a total accuracy of  87.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  3  feature { 7 } was added to current set
Current Set of Features:  [5, 3, 7]  with accuracy of  89.0 %
--------------------------------------------
On level  4  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  87.0 %
--Considering adding feature number  2  gets us a total accuracy of  81.0 %
--Considering adding feature number  4  gets us a total accuracy of  79.0 %
--Considering adding feature number  6  gets us a total accuracy of  88.0 %
--Considering adding feature number  8  gets us a total accuracy of  81.0 %
--Considering adding feature number  9  gets us a total accuracy of  83.0 %
--Considering adding feature number  10  gets us a total accuracy of  84.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  4  feature { 6 } was added to current set
Current Set of Features:  [5, 3, 7, 6]  with accuracy of  88.0 %
--------------------------------------------
On level  5  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  86.0 %
--Considering adding feature number  2  gets us a total accuracy of  83.0 %
--Considering adding feature number  4  gets us a total accuracy of  82.0 %
```

```
--Considering adding feature number  8  gets us a total accuracy of  79.0 %
--Considering adding feature number  9  gets us a total accuracy of  74.0 %
--Considering adding feature number  10  gets us a total accuracy of  80.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  5  feature { 1 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1]  with accuracy of  86.0 %
-------------------------------------------
On level  6  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  73.0 %
--Considering adding feature number  4  gets us a total accuracy of  73.0 %
--Considering adding feature number  8  gets us a total accuracy of  77.0 %
--Considering adding feature number  9  gets us a total accuracy of  70.0 %
--Considering adding feature number  10  gets us a total accuracy of  73.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  6  feature { 8 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1, 8]  with accuracy of  77.0 %
-------------------------------------------
On level  7  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  69.0 %
--Considering adding feature number  4  gets us a total accuracy of  73.0 %
--Considering adding feature number  9  gets us a total accuracy of  71.0 %
--Considering adding feature number  10  gets us a total accuracy of  69.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  7  feature { 4 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1, 8, 4]  with accuracy of  73.0 %
-------------------------------------------
On level  8  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  67.0 %
--Considering adding feature number  9  gets us a total accuracy of  65.0 %
--Considering adding feature number  10  gets us a total accuracy of  70.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  8  feature { 10 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1, 8, 4, 10]  with accuracy of  70.0 %
-------------------------------------------
On level  9  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  67.0 %
```

```
--Considering adding feature number  9  gets us a total accuracy of  62.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  9  feature { 2 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1, 8, 4, 10, 2]  with accuracy of  67.0
%
---------------------------------------------
On level  10  of the search tree

--Considering adding feature number  9  gets us a total accuracy of  68.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  10  feature { 9 } was added to current set
Current Set of Features:  [5, 3, 7, 6, 1, 8, 4, 10, 2, 9]  with accuracy of
68.0 %
---------------------------------------------
Set  [5, 3]  has the highest accuracy: 92.0
Elapsed time: 0.773 seconds
```

%%
%%

Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection Algorithms.

Type in the name of the file to test: **small-test-dataset.txt**

File processed successfully.

Type the number of the algorithm you want to run.

1- Forward Selection.

2- Backward Elimination.

3- Random Combinations.

2

This dataset has 10 features (not including the class attribute), with 100 instances.

Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Running nearest neighbor with all the features, using “leaving-one-out”

evaluation, we get an accuracy of 68.0 %

Beginning search (**Backward Feature Elimination**).

On level 10 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 72.0 %

--Considering removing feature number 9 gets us a total accuracy of 67.0 %

--Considering removing feature number 8 gets us a total accuracy of 72.0 %

--Considering removing feature number 7 gets us a total accuracy of 62.0 %

--Considering removing feature number 6 gets us a total accuracy of 71.0 %

--Considering removing feature number 5 gets us a total accuracy of 69.0 %

--Considering removing feature number 4 gets us a total accuracy of 70.0 %

--Considering removing feature number 3 gets us a total accuracy of 73.0 %

--Considering removing feature number 2 gets us a total accuracy of 62.0 %

--Considering removing feature number 1 gets us a total accuracy of 71.0 %

On level 10 feature { 3 } was removed from current set

Current Set of Features: [1, 2, 4, 5, 6, 7, 8, 9, 10] with accuracy of 73.0 %

On level 9 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 67.0 %

--Considering removing feature number 9 gets us a total accuracy of 73.0 %

--Considering removing feature number 8 gets us a total accuracy of 68.0 %

--Considering removing feature number 7 gets us a total accuracy of 68.0 %

--Considering removing feature number 6 gets us a total accuracy of 75.0 %

--Considering removing feature number 5 gets us a total accuracy of 64.0 %

--Considering removing feature number 4 gets us a total accuracy of 69.0 %

--Considering removing feature number 2 gets us a total accuracy of 73.0 %

--Considering removing feature number 1 gets us a total accuracy of 72.0 %

On level 9 feature { 6 } was removed from current set

Current Set of Features: [1, 2, 4, 5, 7, 8, 9, 10] with accuracy of 75.0 %

On level 8 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 71.0 %

--Considering removing feature number 9 gets us a total accuracy of 77.0 %

--Considering removing feature number 8 gets us a total accuracy of 78.0 %

--Considering removing feature number 7 gets us a total accuracy of 71.0 %

--Considering removing feature number 5 gets us a total accuracy of 57.99999999999999 %

--Considering removing feature number 4 gets us a total accuracy of 64.0 %

--Considering removing feature number 2 gets us a total accuracy of 73.0 %

--Considering removing feature number 1 gets us a total accuracy of 75.0 %

On level 8 feature { 8 } was removed from current set

Current Set of Features: [1, 2, 4, 5, 7, 9, 10] with accuracy of 78.0 %

On level 7 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 76.0 %

--Considering removing feature number 9 gets us a total accuracy of 79.0 %

--Considering removing feature number 7 gets us a total accuracy of 75.0 %

--Considering removing feature number 5 gets us a total accuracy of 61.0 %

--Considering removing feature number 4 gets us a total accuracy of 67.0 %

--Considering removing feature number 2 gets us a total accuracy of 73.0 %

--Considering removing feature number 1 gets us a total accuracy of 76.0 %

On level 7 feature { 9 } was removed from current set

```
Current Set of Features:  [1, 2, 4, 5, 7, 10]  with accuracy of  79.0 %
--------------------------------------------
On level  6  of the search tree

--Considering removing feature number  10  gets us a total accuracy of  77.0 %
--Considering removing feature number  7  gets us a total accuracy of  75.0 %
--Considering removing feature number  5  gets us a total accuracy of  60.0 %
--Considering removing feature number  4  gets us a total accuracy of  75.0 %
--Considering removing feature number  2  gets us a total accuracy of  71.0 %
--Considering removing feature number  1  gets us a total accuracy of  83.0 %

On level  6  feature { 1 } was removed from current set
Current Set of Features:  [2, 4, 5, 7, 10]  with accuracy of  83.0 %
--------------------------------------------
On level  5  of the search tree

--Considering removing feature number  10  gets us a total accuracy of  76.0 %
--Considering removing feature number  7  gets us a total accuracy of  76.0 %
--Considering removing feature number  5  gets us a total accuracy of  71.0 %
--Considering removing feature number  4  gets us a total accuracy of  81.0 %
--Considering removing feature number  2  gets us a total accuracy of  79.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  5  feature { 4 } was removed from current set
Current Set of Features:  [2, 5, 7, 10]  with accuracy of  81.0 %
--------------------------------------------
On level  4  of the search tree

--Considering removing feature number  10  gets us a total accuracy of  77.0 %
--Considering removing feature number  7  gets us a total accuracy of  72.0 %
```

--Considering removing feature number 5 gets us a total accuracy of 68.0 %
--Considering removing feature number 2 gets us a total accuracy of 75.0 %

Warning, Accuracy has decreased! Continuing search in case of local maxima

On level 4 feature { 10 } was removed from current set
Current Set of Features: [2, 5, 7] with accuracy of 77.0 %

On level 3 of the search tree

--Considering removing feature number 7 gets us a total accuracy of 80.0 %
--Considering removing feature number 5 gets us a total accuracy of 54.0 %
--Considering removing feature number 2 gets us a total accuracy of 80.0 %

Warning, Accuracy has decreased! Continuing search in case of local maxima

On level 3 feature { 7 } was removed from current set
Current Set of Features: [2, 5] with accuracy of 80.0 %

On level 2 of the search tree

--Considering removing feature number 5 gets us a total accuracy of 54.0 %
--Considering removing feature number 2 gets us a total accuracy of 75.0 %

Warning, Accuracy has decreased! Continuing search in case of local maxima

On level 2 feature { 2 } was removed from current set
Current Set of Features: [5] with accuracy of 75.0 %

On level 1 of the search tree

Current Set of Features: [] with accuracy of 75.0 %

Set [2, 4, 5, 7, 10] has the highest accuracy: 83.0

Elapsed time: 1.077 seconds

%%

Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection Algorithms.

Type in the name of the file to test: **small-test-dataset.txt**

File processed successfully.

Type the number of the algorithm you want to run.

1- Forward Selection.

2- Backward Elimination.

3- Random Combinations.

3

This dataset has 10 features (not including the class attribute), with 100 instances.

Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Default number of subsets is 10 and number of features in each set is log(10)= 3

Do you want to change the default values? (y/n): n

Num of total features: 10 Features number in one set: 3

Features sets: [(1, 2, 6), (3, 4, 10), (2, 3, 10), (3, 5, 8), (8, 9, 10), (1, 8, 10), (5, 6, 10), (4, 6, 9), (2, 3, 4), (2, 4, 5)]

Beginning search (**Random Feature Combinations**).

```
---------------------------------------------

Accuracy for feature set (1, 2, 6) : 64.0

Accuracy for feature set (3, 4, 10) : 61.0

Accuracy for feature set (2, 3, 10) : 72.0

Accuracy for feature set (3, 5, 8) : 79.0

Accuracy for feature set (8, 9, 10) : 65.0

Accuracy for feature set (1, 8, 10) : 62.0

Accuracy for feature set (5, 6, 10) : 81.0

Accuracy for feature set (4, 6, 9) : 59.0

Accuracy for feature set (2, 3, 4) : 70.0

Accuracy for feature set (2, 4, 5) : 74.0


Set  (5, 6, 10)  has the highest accuracy: 81.0

Elapsed time: 0.049 seconds


Process finished with exit code 0
```

%%%
%%

```
Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection
Algorithms.
Type in the name of the file to test: CS170_Spring_2023_Small_data__44.txt
File processed successfully.
Type the number of the algorithm you want to run.
1- Forward Selection.
2- Backward Elimination.
3- Random Combinations.
1

This dataset has 10 features (not including the class attribute), with 100
instances.
Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Running nearest neighbor with no features (default rate), using
“leaving-one-out” evaluation, we get an accuracy of  86.0 %
---------------------------------------------
```

Beginning search (**Forward Feature Selection**).

On level 1 of the search tree

--Considering adding feature number 1 gets us a total accuracy of 73.0 %
--Considering adding feature number 2 gets us a total accuracy of 77.0 %
--Considering adding feature number 3 gets us a total accuracy of 88.0 %
--Considering adding feature number 4 gets us a total accuracy of 76.0 %
--Considering adding feature number 5 gets us a total accuracy of 78.0 %
--Considering adding feature number 6 gets us a total accuracy of 79.0 %
--Considering adding feature number 7 gets us a total accuracy of 78.0 %
--Considering adding feature number 8 gets us a total accuracy of 72.0 %
--Considering adding feature number 9 gets us a total accuracy of 75.0 %
--Considering adding feature number 10 gets us a total accuracy of 74.0 %

On level 1 feature { 3 } was added to current set
Current Set of Features: [3] with accuracy of 88.0 %

On level 2 of the search tree

--Considering adding feature number 1 gets us a total accuracy of 84.0 %
--Considering adding feature number 2 gets us a total accuracy of 80.0 %
--Considering adding feature number 4 gets us a total accuracy of 84.0 %
--Considering adding feature number 5 gets us a total accuracy of 95.0 %
--Considering adding feature number 6 gets us a total accuracy of 82.0 %
--Considering adding feature number 7 gets us a total accuracy of 90.0 %
--Considering adding feature number 8 gets us a total accuracy of 84.0 %
--Considering adding feature number 9 gets us a total accuracy of 76.0 %
--Considering adding feature number 10 gets us a total accuracy of 88.0 %

On level 2 feature { 5 } was added to current set
Current Set of Features: [3, 5] with accuracy of 95.0 %

On level 3 of the search tree

--Considering adding feature number 1 gets us a total accuracy of 86.0 %
--Considering adding feature number 2 gets us a total accuracy of 84.0 %
--Considering adding feature number 4 gets us a total accuracy of 84.0 %
--Considering adding feature number 6 gets us a total accuracy of 86.0 %
--Considering adding feature number 7 gets us a total accuracy of 90.0 %
--Considering adding feature number 8 gets us a total accuracy of 84.0 %
--Considering adding feature number 9 gets us a total accuracy of 77.0 %
--Considering adding feature number 10 gets us a total accuracy of 92.0 %

Warning, Accuracy has decreased! Continuing search in case of local maxima

On level 3 feature { 10 } was added to current set

```
Current Set of Features:  [3, 5, 10]  with accuracy of  92.0 %
-------------------------------------------
On level  4  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  86.0 %
--Considering adding feature number  2  gets us a total accuracy of  76.0 %
--Considering adding feature number  4  gets us a total accuracy of  84.0 %
--Considering adding feature number  6  gets us a total accuracy of  76.0 %
--Considering adding feature number  7  gets us a total accuracy of  85.0 %
--Considering adding feature number  8  gets us a total accuracy of  80.0 %
--Considering adding feature number  9  gets us a total accuracy of  87.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  4  feature { 9 } was added to current set
Current Set of Features:  [3, 5, 10, 9]  with accuracy of  87.0 %
-------------------------------------------
On level  5  of the search tree

--Considering adding feature number  1  gets us a total accuracy of  82.0 %
--Considering adding feature number  2  gets us a total accuracy of  82.0 %
--Considering adding feature number  4  gets us a total accuracy of  82.0 %
--Considering adding feature number  6  gets us a total accuracy of  76.0 %
--Considering adding feature number  7  gets us a total accuracy of  82.0 %
--Considering adding feature number  8  gets us a total accuracy of  75.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  5  feature { 1 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1]  with accuracy of  82.0 %
-------------------------------------------
On level  6  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  80.0 %
--Considering adding feature number  4  gets us a total accuracy of  79.0 %
--Considering adding feature number  6  gets us a total accuracy of  84.0 %
--Considering adding feature number  7  gets us a total accuracy of  76.0 %
--Considering adding feature number  8  gets us a total accuracy of  77.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  6  feature { 6 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1, 6]  with accuracy of  84.0 %
-------------------------------------------
On level  7  of the search tree

--Considering adding feature number  2  gets us a total accuracy of  83.0 %
--Considering adding feature number  4  gets us a total accuracy of  78.0 %
```

```
--Considering adding feature number  7  gets us a total accuracy of  77.0 %
--Considering adding feature number  8  gets us a total accuracy of  79.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  7  feature { 2 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1, 6, 2]  with accuracy of  83.0 %
---------------------------------------------
On level  8  of the search tree

--Considering adding feature number  4  gets us a total accuracy of  76.0 %
--Considering adding feature number  7  gets us a total accuracy of  74.0 %
--Considering adding feature number  8  gets us a total accuracy of  77.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  8  feature { 8 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1, 6, 2, 8]  with accuracy of  77.0 %
---------------------------------------------
On level  9  of the search tree

--Considering adding feature number  4  gets us a total accuracy of  77.0 %
--Considering adding feature number  7  gets us a total accuracy of  77.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  9  feature { 4 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1, 6, 2, 8, 4]  with accuracy of  77.0
%
---------------------------------------------
On level  10  of the search tree

--Considering adding feature number  7  gets us a total accuracy of  72.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  10  feature { 7 } was added to current set
Current Set of Features:  [3, 5, 10, 9, 1, 6, 2, 8, 4, 7]  with accuracy of
72.0 %
---------------------------------------------
Set  [3, 5]  has the highest accuracy: 95.0
Elapsed time: 0.791 seconds

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection
Algorithms.
```

Type in the name of the file to test: **CS170_Spring_2023_Small_data__44.txt**
File processed successfully.
Type the number of the algorithm you want to run.
1- Forward Selection.
2- Backward Elimination.
3- Random Combinations.
2

This dataset has 10 features (not including the class attribute), with 100 instances.
Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Running nearest neighbor with all the features, using “leaving-one-out” evaluation, we get an accuracy of 72.0 %

Beginning search (**Backward Feature Elimination**).

On level 10 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 76.0 %
--Considering removing feature number 9 gets us a total accuracy of 75.0 %
--Considering removing feature number 8 gets us a total accuracy of 74.0 %
--Considering removing feature number 7 gets us a total accuracy of 77.0 %
--Considering removing feature number 6 gets us a total accuracy of 78.0 %
--Considering removing feature number 5 gets us a total accuracy of 70.0 %
--Considering removing feature number 4 gets us a total accuracy of 77.0 %
--Considering removing feature number 3 gets us a total accuracy of 69.0 %
--Considering removing feature number 2 gets us a total accuracy of 71.0 %
--Considering removing feature number 1 gets us a total accuracy of 66.0 %

On level 10 feature { 6 } was removed from current set
Current Set of Features: [1, 2, 3, 4, 5, 7, 8, 9, 10] with accuracy of 78.0 %

On level 9 of the search tree

--Considering removing feature number 10 gets us a total accuracy of 79.0 %
--Considering removing feature number 9 gets us a total accuracy of 82.0 %
--Considering removing feature number 8 gets us a total accuracy of 78.0 %
--Considering removing feature number 7 gets us a total accuracy of 75.0 %
--Considering removing feature number 5 gets us a total accuracy of 74.0 %
--Considering removing feature number 4 gets us a total accuracy of 77.0 %
--Considering removing feature number 3 gets us a total accuracy of 72.0 %
--Considering removing feature number 2 gets us a total accuracy of 76.0 %
--Considering removing feature number 1 gets us a total accuracy of 73.0 %

```
On level  9  feature { 9 } was removed from current set
Current Set of Features:  [1, 2, 3, 4, 5, 7, 8, 10]  with accuracy of  82.0 %
--------------------------------------------
On level  8  of the search tree

--Considering removing feature number  10  gets us a total accuracy of  82.0 %
--Considering removing feature number  8  gets us a total accuracy of  85.0 %
--Considering removing feature number  7  gets us a total accuracy of  73.0 %
--Considering removing feature number  5  gets us a total accuracy of  76.0 %
--Considering removing feature number  4  gets us a total accuracy of  83.0 %
--Considering removing feature number  3  gets us a total accuracy of  76.0 %
--Considering removing feature number  2  gets us a total accuracy of  80.0 %
--Considering removing feature number  1  gets us a total accuracy of  76.0 %

On level  8  feature { 8 } was removed from current set
Current Set of Features:  [1, 2, 3, 4, 5, 7, 10]  with accuracy of  85.0 %
--------------------------------------------
On level  7  of the search tree

--Considering removing feature number  10  gets us a total accuracy of  84.0 %
--Considering removing feature number  7  gets us a total accuracy of  78.0 %
--Considering removing feature number  5  gets us a total accuracy of  80.0 %
--Considering removing feature number  4  gets us a total accuracy of  82.0 %
--Considering removing feature number  3  gets us a total accuracy of  81.0 %
--Considering removing feature number  2  gets us a total accuracy of  80.0 %
--Considering removing feature number  1  gets us a total accuracy of  78.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  7  feature { 10 } was removed from current set
Current Set of Features:  [1, 2, 3, 4, 5, 7]  with accuracy of  84.0 %
--------------------------------------------
On level  6  of the search tree

--Considering removing feature number  7  gets us a total accuracy of  78.0 %
--Considering removing feature number  5  gets us a total accuracy of  83.0 %
--Considering removing feature number  4  gets us a total accuracy of  89.0 %
--Considering removing feature number  3  gets us a total accuracy of  83.0 %
--Considering removing feature number  2  gets us a total accuracy of  82.0 %
--Considering removing feature number  1  gets us a total accuracy of  81.0 %

On level  6  feature { 4 } was removed from current set
Current Set of Features:  [1, 2, 3, 5, 7]  with accuracy of  89.0 %
--------------------------------------------
On level  5  of the search tree

--Considering removing feature number  7  gets us a total accuracy of  81.0 %
```

```
--Considering removing feature number  5  gets us a total accuracy of  90.0 %
--Considering removing feature number  3  gets us a total accuracy of  83.0 %
--Considering removing feature number  2  gets us a total accuracy of  83.0 %
--Considering removing feature number  1  gets us a total accuracy of  86.0 %

On level  5  feature { 5 } was removed from current set
Current Set of Features:  [1, 2, 3, 7]  with accuracy of  90.0 %
--------------------------------------------
On level  4  of the search tree

--Considering removing feature number  7  gets us a total accuracy of  80.0 %
--Considering removing feature number  3  gets us a total accuracy of  84.0 %
--Considering removing feature number  2  gets us a total accuracy of  84.0 %
--Considering removing feature number  1  gets us a total accuracy of  84.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  4  feature { 3 } was removed from current set
Current Set of Features:  [1, 2, 7]  with accuracy of  84.0 %
--------------------------------------------
On level  3  of the search tree

--Considering removing feature number  7  gets us a total accuracy of  81.0 %
--Considering removing feature number  2  gets us a total accuracy of  77.0 %
--Considering removing feature number  1  gets us a total accuracy of  79.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  3  feature { 7 } was removed from current set
Current Set of Features:  [1, 2]  with accuracy of  81.0 %
--------------------------------------------
On level  2  of the search tree

--Considering removing feature number  2  gets us a total accuracy of  73.0 %
--Considering removing feature number  1  gets us a total accuracy of  77.0 %

**Warning, Accuracy has decreased! Continuing search in case of local maxima**

On level  2  feature { 1 } was removed from current set
Current Set of Features:  [2]  with accuracy of  77.0 %
--------------------------------------------
On level  1  of the search tree

Current Set of Features: [] with accuracy of  86.0 %
--------------------------------------------
Set  [1, 2, 3, 7]  has the highest accuracy: 90.0
Elapsed time: 1.126 seconds
```

%%
%%

Welcome to Abhinav's, Majd's, Jacqueline's and Rovin's Feature Selection Algorithms.
Type in the name of the file to test: **CS170_Spring_2023_Small_data__44.txt**
File processed successfully.
Type the number of the algorithm you want to run.
1- Forward Selection.
2- Backward Elimination.
3- Random Combinations.
3

This dataset has 10 features (not including the class attribute), with 100 instances.
Please wait while I normalize the data...

Do you want to Z Normalize the data values? (y/n): n

Default number of subsets is 10 and number of features in each set is log(10)= 3
Do you want to change the default values? (y/n): n

Num of total features: 10 Features number in one set: 3

Features sets: [(4, 5, 6), (4, 7, 8), (7, 8, 9), (3, 5, 6), (1, 5, 6), (1, 3, 7), (1, 6, 8), (7, 9, 10), (1, 3, 4), (2, 6, 10)]

Beginning search (**Random Feature Combinations**).

Accuracy for feature set (4, 5, 6) : 72.0
Accuracy for feature set (4, 7, 8) : 72.0
Accuracy for feature set (7, 8, 9) : 73.0
Accuracy for feature set (3, 5, 6) : 86.0
Accuracy for feature set (1, 5, 6) : 74.0
Accuracy for feature set (1, 3, 7) : 84.0
Accuracy for feature set (1, 6, 8) : 76.0
Accuracy for feature set (7, 9, 10) : 76.0
Accuracy for feature set (1, 3, 4) : 80.0
Accuracy for feature set (2, 6, 10) : 76.0

Set (3, 5, 6) has the highest accuracy: 86.0
Elapsed time: 0.047 seconds

Process finished with exit code 0

%%
%%