

**Luis Sanchez**

SID# 862046663

Email: [Lsanc044@ucr.edu](mailto:Lsanc044@ucr.edu)

March 18th, 2021

Project 2 for CS170 Winter 2021, with Dr. Eamonn Keogh

All important code is original excluding c++ standard libraries

- `<iostream>`
- `<fstream>`
- `<sstream>`
- `<string>`
- `<vector>`
- `<cstdio>`
- `<cmath>`

For my second project in CS170 Artificial intelligence I was tasked by Dr. Keogh to create a feature selection with the nearest neighbor algorithm. Dr. K introduced us to the nearest neighbor algorithm in lecture and provided a small and large data set to test on.

I first ran forward selection on the small data set and the results can be seen in Figure 1. The small file that was assigned to me was CS170\_SMALLtestdata\_\_69.txt.

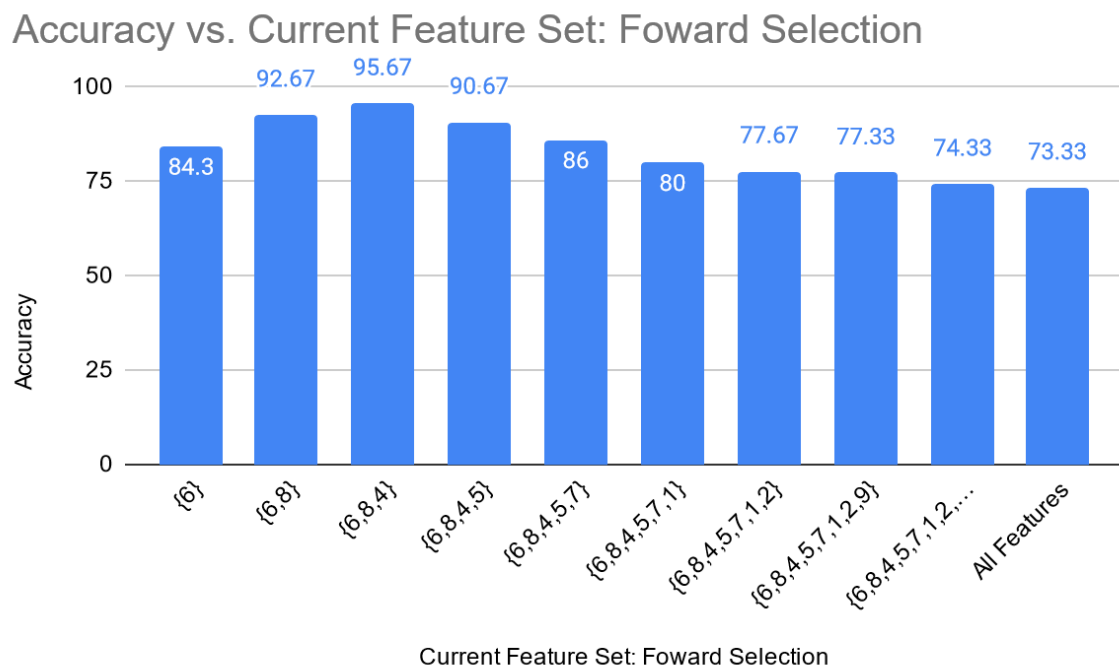


Figure 1: Forward selection on small data set

At the start of the search the first feature that was chosen was the 6th feature with an accuracy of 84.3%. The algorithm then added features 8 and 4 for an accuracy improvement to 92.67 and 95.67 respectively. After the 3rd level the accuracy started to go down but I kept running the algorithm till all features were chosen and the accuracy continued to go down. The best features to use for my small data set was 6,8,4 with an accuracy of 95.67%

Next, as shown in Figure 2, I ran a backwards elimination of the same small set.

### Accuracy vs. Current Feature Set: Backward Elimination

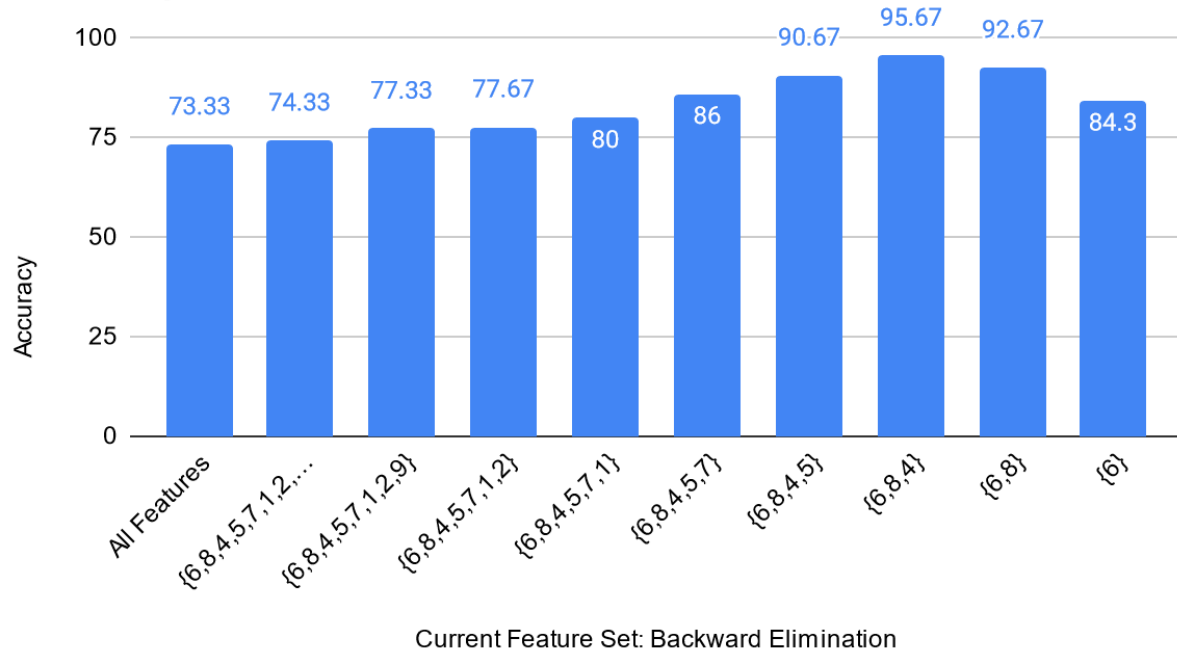


Figure 2: Backwards elimination on small dataset

The end result of a backwards elimination is very similar to the forward elimination. With all features I got the same accuracy of 73.33 and from there the algorithm removes the worst feature till all features have been eliminated. The algorithm continued slowly increasing in accuracy till hitting 5 elements and jumping up 6% then followed by a 4% improvement till reaching another 5% improvement with 3 elements {6,8,4}, the same elements from the forward selection.

**Conclusion for Small Dataset:** From my testing I found that features {6,8,4} to best the most accurate with an accuracy of 95.67% for small dataset #69

For the large dataset I was assigned CS170\_targettestdata\_\_59.txt. Figure 3 displays my findings from forward selection on the large dataset.

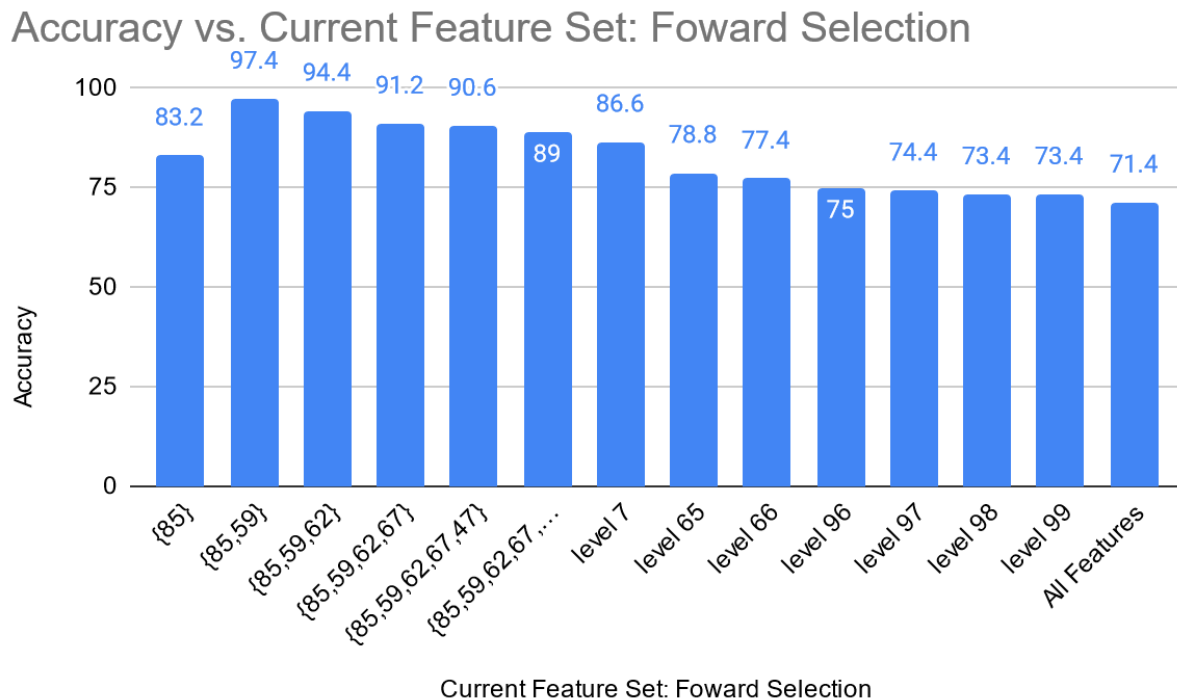


Figure 3: Forward selection on small data set

For figure 3 I omitted some middle level of the tree to save space. The best accuracy I found was with elements {85,59} with an accuracy of 97.4%. Farther down the tree the accuracy slowly decreased. I added a middle snapshot around level 65-66 that showed an accuracy of 78.8 and 77.4 respectfully. Then for the last 4 levels the accuracy went from 74.4 down to 71.4 with all features.

**Conclusion on the large data set:** I found the features 85 and 59 to be the most accurate with an accuracy of 97.4 for large data set # 59

**Computational effort for search:** I ran all my experiments on the ucr bolt servers. In table 1 I report the run time for the searches I ran.

	Small(10 features, 300 instances)	Large(100 features, 500 instances)
Forward	4.68 seconds	54 minutes
Backward	4.82 seconds	

**Print out of my C++ code: [Github](#)**

**Single Trace of algorithm with forward selection on small data set:**

Welcome to Luis Sanchez's Feature Selection Algorithm.

Type in the name of the file to test :Type the number of the algorithm you want to run.

1)Forward Selection

2)Backward Elimination

1

This dataset has 10 features (not including the class attribute), with 300 instances.

Beginning search.

On the 1th level of the search tree

- accuracy for feature {1} 65
- accuracy for feature {2} 64.3333
- accuracy for feature {3} 65
- accuracy for feature {4} 67.6667
- accuracy for feature {5} 65
- accuracy for feature {6} 84.3333
- accuracy for feature {7} 64.3333
- accuracy for feature {8} 70
- accuracy for feature {9} 67
- accuracy for feature {10} 66.6667

On level 1 I added feature 6 to the best features set

On the 2th level of the search tree

- accuracy for features {6,1} 85.6667%
- accuracy for features {6,2} 82.3333%
- accuracy for features {6,3} 77%
- accuracy for features {6,4} 83%
- accuracy for features {6,5} 84.6667%
- accuracy for features {6,7} 83%
- accuracy for features {6,8} 92.6667%
- accuracy for features {6,9} 80.6667%
- accuracy for features {6,10} 83.3333%

On level 2 I added feature 8 to the best features set

On the 3th level of the search tree

- accuracy for features {6,8,1} 91.3333%
- accuracy for features {6,8,2} 88%
- accuracy for features {6,8,3} 86.3333%
- accuracy for features {6,8,4} 95.6667%
- accuracy for features {6,8,5} 90%
- accuracy for features {6,8,7} 87%
- accuracy for features {6,8,9} 89.3333%
- accuracy for features {6,8,10} 87%

On level 3 I added feature 4 to the best features set

On the 4th level of the search tree

- accuracy for features {6,8,4,1} 90%
- accuracy for features {6,8,4,2} 88.6667%
- accuracy for features {6,8,4,3} 86.6667%
- accuracy for features {6,8,4,5} 90.6667%
- accuracy for features {6,8,4,7} 90.3333%
- accuracy for features {6,8,4,9} 87.3333%
- accuracy for features {6,8,4,10} 90%

On level 4 I added feature 5 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 5th level of the search tree

- accuracy for features {6,8,4,5,1} 83.6667%
- accuracy for features {6,8,4,5,2} 83.6667%
- accuracy for features {6,8,4,5,3} 83.3333%
- accuracy for features {6,8,4,5,7} 86%
- accuracy for features {6,8,4,5,9} 83%
- accuracy for features {6,8,4,5,10} 83.3333%

On level 5 I added feature 7 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 6th level of the search tree

- accuracy for features {6,8,4,5,7,1} 80%
- accuracy for features {6,8,4,5,7,2} 79.6667%
- accuracy for features {6,8,4,5,7,3} 80%
- accuracy for features {6,8,4,5,7,9} 78.3333%
- accuracy for features {6,8,4,5,7,10} 78.3333%

On level 6 I added feature 1 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 7th level of the search tree

- accuracy for features {6,8,4,5,7,1,2} 77.6667%
- accuracy for features {6,8,4,5,7,1,3} 77%
- accuracy for features {6,8,4,5,7,1,9} 77.3333%
- accuracy for features {6,8,4,5,7,1,10} 72.6667%

On level 7 I added feature 2 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 8th level of the search tree

- accuracy for features {6,8,4,5,7,1,2,3} 74%
- accuracy for features {6,8,4,5,7,1,2,9} 76.3333%
- accuracy for features {6,8,4,5,7,1,2,10} 73%

On level 8 I added feature 9 to the best features set



(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 9th level of the search tree  
accuracy for features {6,8,4,5,7,1,2,9,3} 74.3333%  
accuracy for features {6,8,4,5,7,1,2,9,10} 72%

On level 9 I added feature 3 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

On the 10th level of the search tree  
accuracy for features {6,8,4,5,7,1,2,9,3,10} 73.3333%

On level 10 I added feature 10 to the best features set

(Warning, Accuracy has decreased! Continuing search in case of local maxima)  
Feature set {6,8,4,} 95.6667%

Finished search!! The best  
Feature set {6,8,4,} with an accuracy of 95.6667%