Project 2: Feature Selection with Nearest Neighbor

Student Name1: SID: Lecture Session:

Student Name2: SID: Lecture Session:

Student Name3: SID: Lecture Session:

Student Name4: SID: Lecture Session:

Solution: <the datasets are your uniquely assigned datasets>

Dataset	Best Feature Set	Accuracy
Small Number: <insert your small dataset number>	Forward Selection = {3, 4 ,5}	0.89
	Backward Elimination = {3, 5 7}	0.78
	Custom Algorithm = Not implemented	0.91
Large Number: <insert your large dataset number>	Forward Selection = {15, 4 ,1}	0.81
	Backward Elimination = {15, 16, 7}	0.92
	Custom Algorithm = {15, 4, 3}	0.87

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

In completing this project, I consulted following resources:

<mention your resources here>

Contribution of each student in the group:

I. Introduction

II. Challenges

III. Code Design

<code documentation>

IV. Dataset details

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

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

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

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

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

V. Algorithms

1. Forward Selection
2. Backward Elimination
3. Your custom algorithm (optional)

<explain each algorithm in brief>

VI. Analysis

Experiment 1: Comparing Forward Selection vs Backward Elimination.

<do this experiment for sure>

Compare accuracy with no feature selection vs with feature selection.

Compare feature set and accuracy for forward selection vs backward elimination.

Talk about pros and cons of both algorithms.

Experiment 2: Effect of normalization

Compare accuracy when using normalized vs unnormalized data.

Experiment 3: Effect of number neighbors (k)

Plot accuracy vs increasing values of k and examine the trend.

VII. Conclusion

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

VIII. Trace of your small dataset

<paste the trace of your small dataset here>