UNIVERSITI MALAYA

Faculty of Computer Science and Information Technology

WIX3001

SOFT COMPUTING

OCC 3

ASSIGNMENT 1 MATLAB PROGRAMMING

LECTURER NAME: DR. LIEW WEI SHIUNG

STUDENT NAME: LIM HON TING

MATRIC NUMBER: S2114212

DATE: 29th APRIL 2023

Table of Content

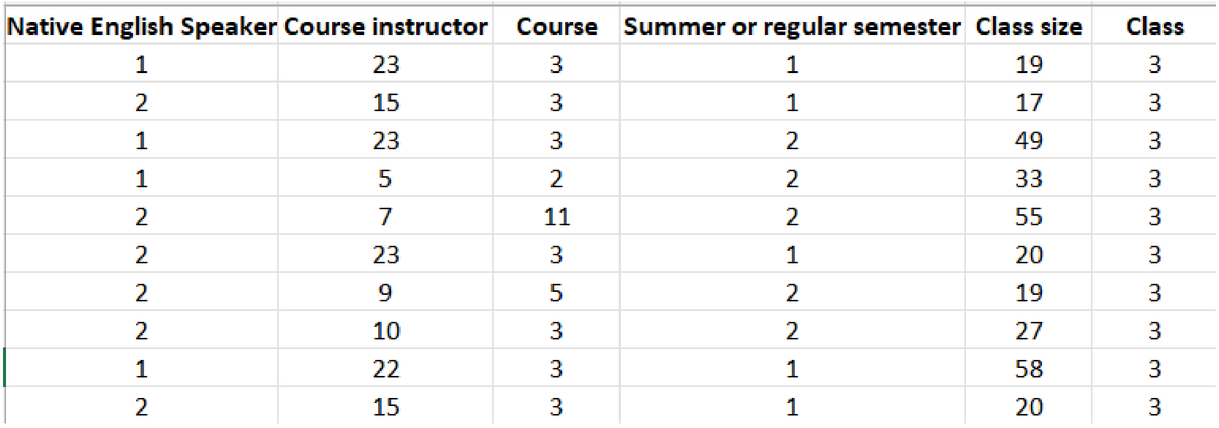
1 Teaching Assistant Evaluation Dataset

1.1 Introduction

The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into 3 roughly equal-sized categories ("low", "medium", and "high") to form the class variable.

1.2 Samples

The dataset contains a total of 151 samples, with each sample representing an evaluation for a teaching assistant from one of the three different categories.



Native English Speaker	Course instructor	Course	Summer or regular semester	Class size	Class
1	23	3	1	19	3
2	15	3	1	17	3
1	23	3	2	49	3
1	5	2	2	33	3
2	7	11	2	55	3
2	23	3	1	20	3
2	9	5	2	19	3
2	10	3	2	27	3
1	22	3	1	58	3
2	15	3	1	20	3

1.3 Features

The dataset consists of 5 attributes, which includes the course information, class size and proficiency with English. The attributes included in the dataset are as follows:

1. Native english speaker
2. Course instructor
3. Course
4. Summer or regular semester
5. Class size

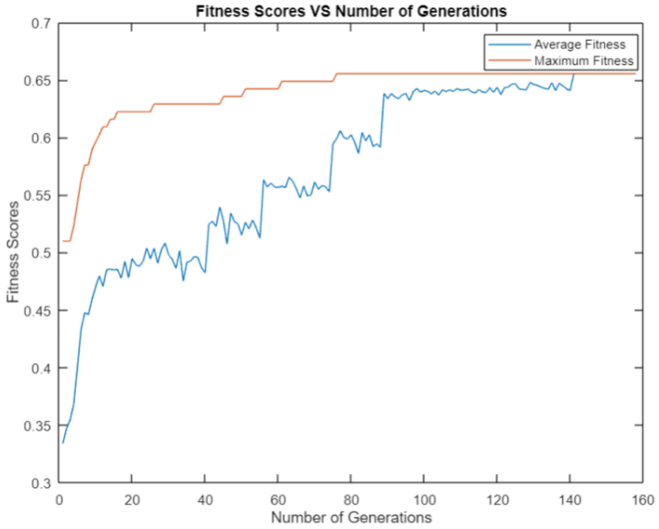
1.4 Classes

The dataset is a three categories classification task, with the classes being low, medium and high scores, at the end of the normal duration of the course. The class distribution is as follows:

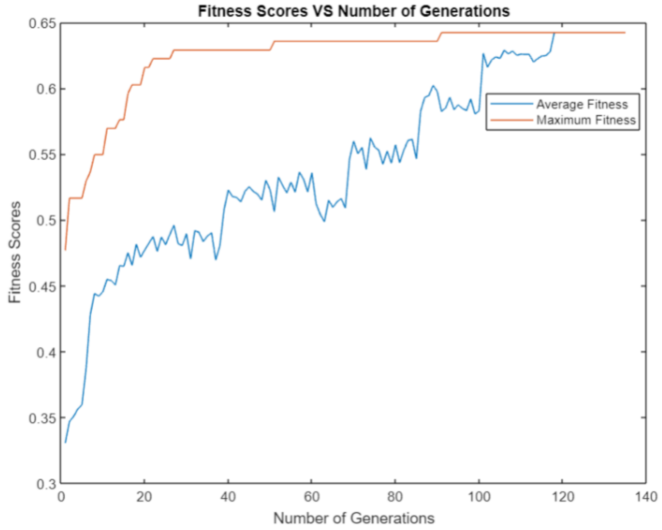
Class 1 - Low: 49 instances
Class 2 - Medium: 50 instances
Class 3 - High: 52 instances

1.5 Results

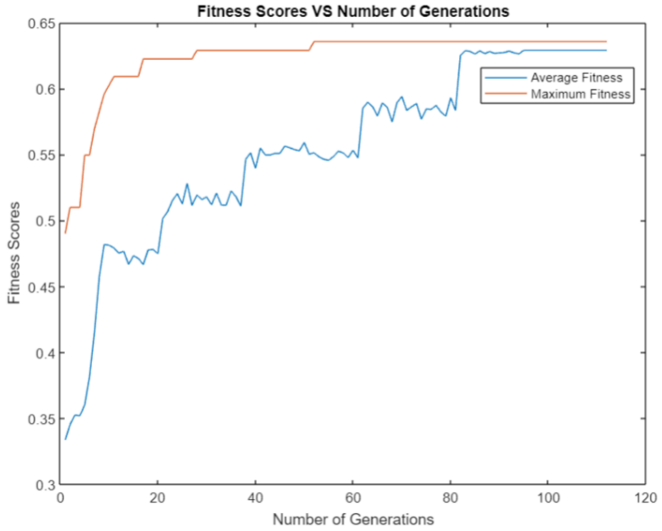
	Generation =1		Generation = Last Generation			
SeedNumber	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for each Layer
723	0.33377	0.50993	0.65563	0.65563	**Avg**: 1 **Std**: 0	**Avg**: 10 **Std**: 0
1075	0.3304	0.47682	0.64238	0.64238	**Avg**: 1 **Std**: 0	**Avg**: 10 **Std**: 0
957	0.33347	0.49007	0.62921	0.63576	**Avg**: 1 **Std**: 0	**Avg**: 4 **Std**: 0



Graph 1.1 - Graph for Teaching Assistant Evaluation Dataset Seed Number 723



Graph 1.2 - Graph for Teaching Assistant Evaluation Dataset Seed Number 1075



Graph 1.3 - Graph for Teaching Assistant Evaluation Dataset Seed Number 957

1.6 Resources

The dataset is available online at

https://archive.ics.uci.edu/ml/datasets/teaching+assistant+evaluation

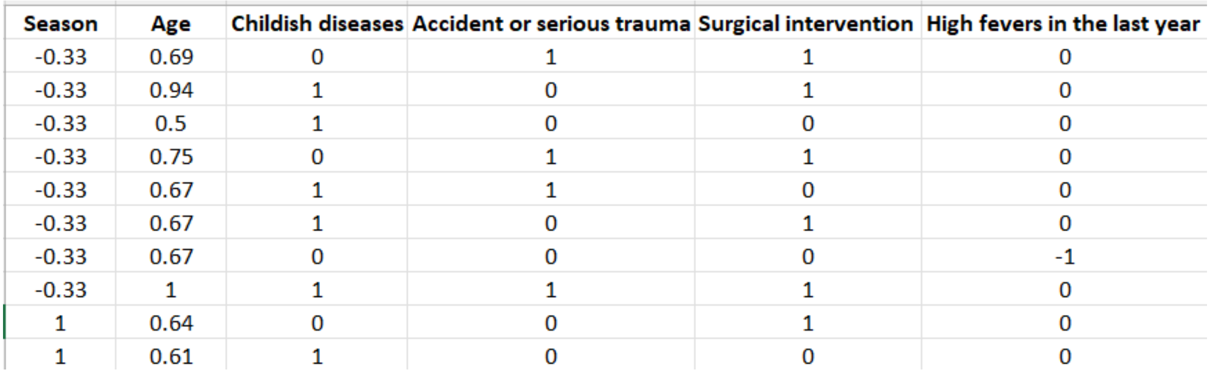
2 Fertility Diagnosis Dataset

2.1 Introduction

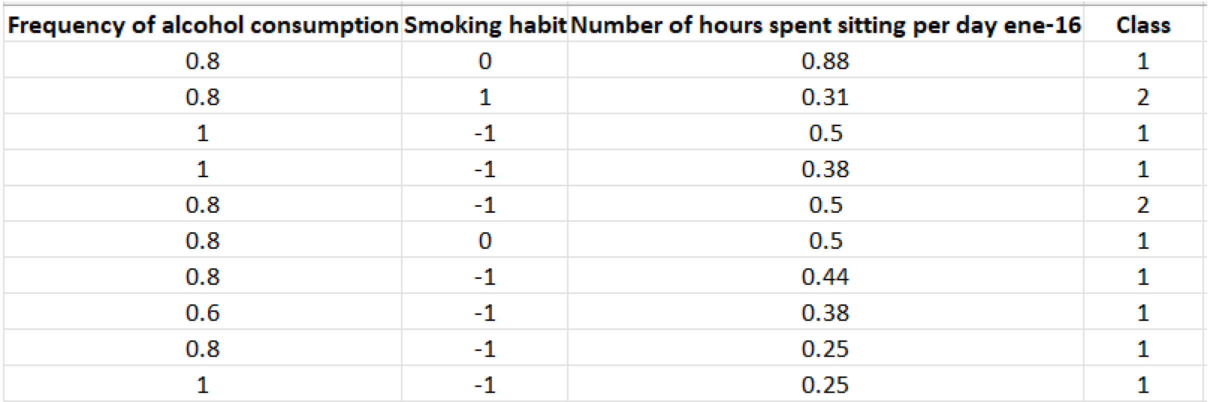
100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits.

2.2 Samples

The dataset contains a total of 100 samples, with each sample will be classified as normal or altered.



Season	Age	Childish diseases	Accident or serious trauma	Surgical intervention	High fevers in the last year
-0.33	0.69	0	1	1	0
-0.33	0.94	1	0	1	0
-0.33	0.5	1	0	0	0
-0.33	0.75	0	1	1	0
-0.33	0.67	1	1	0	0
-0.33	0.67	1	0	1	0
-0.33	0.67	0	0	0	-1
-0.33	1	1	1	1	0
1	0.64	0	0	1	0
1	0.61	1	0	0	0



Frequency of alcohol consumption	Smoking habit	Number of hours spent sitting per day ene-16	Class
0.8	0	0.88	1
0.8	1	0.31	2
1	-1	0.5	1
1	-1	0.38	1
0.8	-1	0.5	2
0.8	0	0.5	1
0.8	-1	0.44	1
0.6	-1	0.38	1
0.8	-1	0.25	1
1	-1	0.25	1

2.3 Features

The dataset consists of 9 attributes, which include their life habits, health status and environmental data. The attributes included in the dataset are as follows:

1. Season in which the analysis was performed.

2. Age at the time of analysis.
3. Childish diseases
4. Accident or serious trauma
5. Surgical intervention
6. High fevers in the last year
7. Frequency of alcohol consumption
8. Smoking habit
9. Number of hours spent sitting per day ene-16

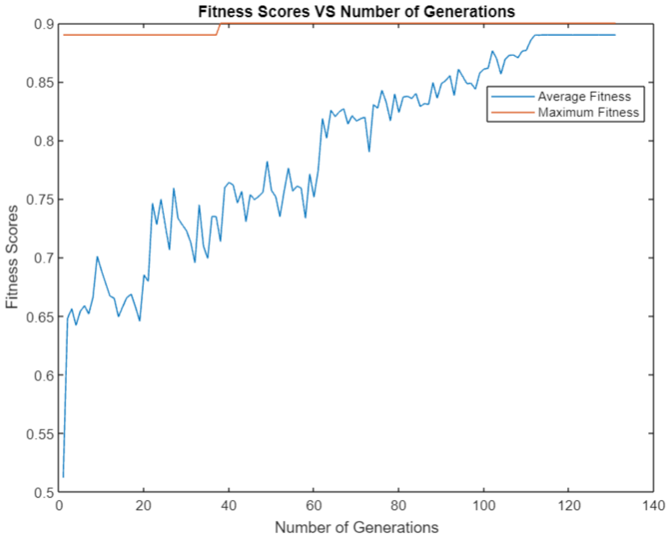
2.4 Classes

The dataset is divided into two classes, which represent normal or altered. The class distribution is as follows:

Class 1: Normal - 88 instances
Class 2: Altered - 12 instances

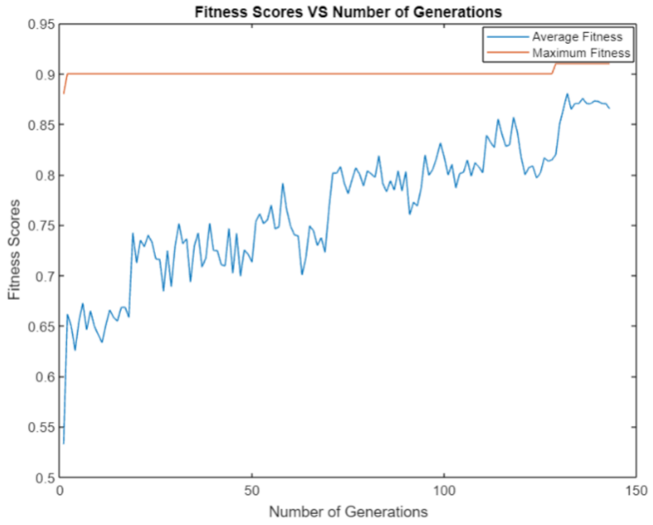
2.5 Results

	Generation =1		Generation = Last Generation			
SeedNumber	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for each Layer
507	0.51183	0.89	0.89007	0.9	**Avg**: 1 **Std**: 0	**Avg**: 6.997 **Std**: 0.058
960	0.5327	0.88	0.8806	0.91	**Avg**: 3.873 **Std**: 2.112	**Avg**: 5.39, 5.5415, 5.157, 5.975, 5.074, 5.133, 6.211, 5.278, 3.714, 5.5 **Std**: 2.11, 1.741, 1.921, 2.093, 1.985, 2.172, 2.117, 1.85 1.9795 ,2.598
1068	0.50163	0.88	0.91	0.91	**Avg**: 1 **Std**: 0	**Avg**: 7 **Std**: 0



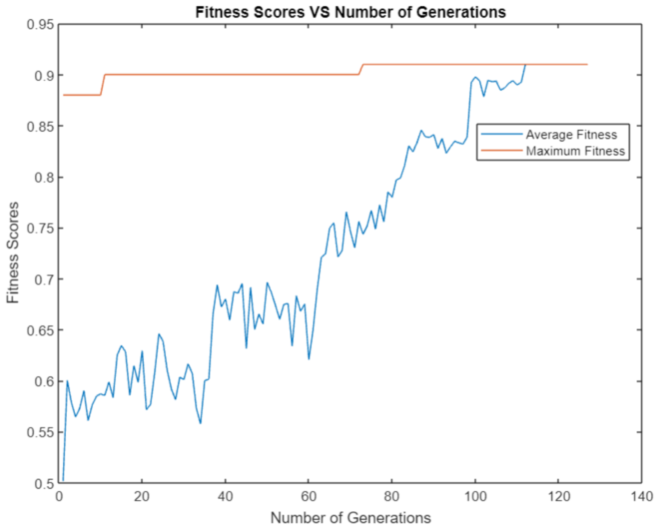
Graph 2.1 - Graph for Fertility Diagnosis Dataset

Seed Number 507



Graph 2.2 - Graph for Fertility Diagnosis Dataset

Seed Number 960



Graph 2.3 - Graph for Fertility Diagnosis Dataset

Seed Number 1068

2.6 Resources

The dataset is available online at

https://archive.ics.uci.edu/ml/datasets/Fertility

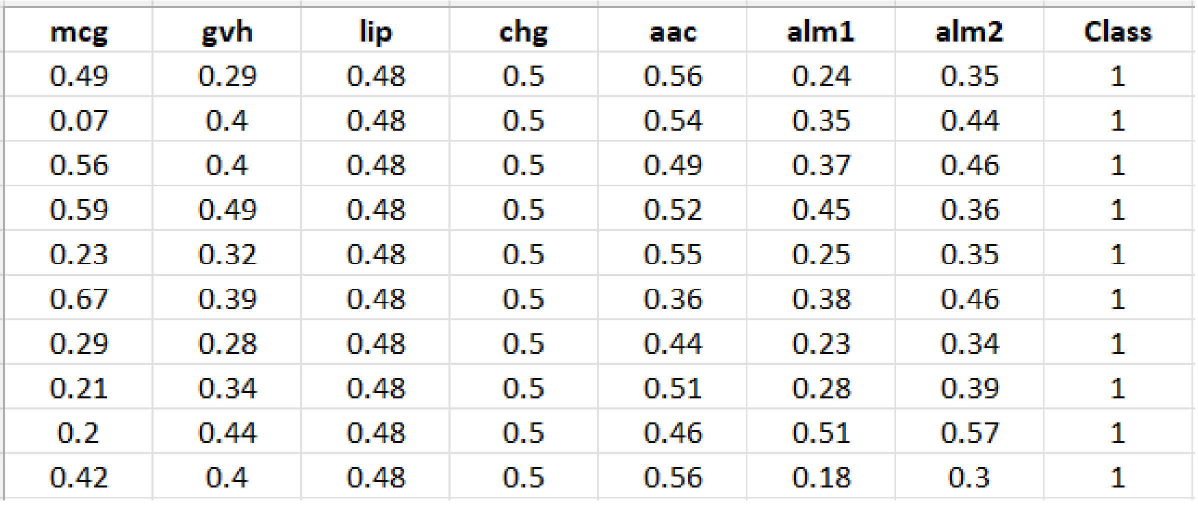
3 Ecoli Dataset

3.1 Introduction

The dataset provides information and attributes that can be used to predict whether a protein is localized in the cytoplasmic or periplasmic region, or the inner membrane of the Ecoli bacteria. The Ecoli dataset includes information about the subcellular localization sites of proteins in E. coli, specifically eight different localization sites which are cytoplasm, inner membrane without signal sequence, perisplasm, inner membrane, uncleavable signal sequence, outer membrane, outer membrane lipoprotein, inner membrane lipoprotein, and inner membrane, cleavable signal sequence.

3.2 Samples

The dataset contains a total of 336 samples, with each sample representing a protein site from one of the eight different protein localization sites.



mcg	gvh	lip	chg	aac	alm1	alm2	Class
0.49	0.29	0.48	0.5	0.56	0.24	0.35	1
0.07	0.4	0.48	0.5	0.54	0.35	0.44	1
0.56	0.4	0.48	0.5	0.49	0.37	0.46	1
0.59	0.49	0.48	0.5	0.52	0.45	0.36	1
0.23	0.32	0.48	0.5	0.55	0.25	0.35	1
0.67	0.39	0.48	0.5	0.36	0.38	0.46	1
0.29	0.28	0.48	0.5	0.44	0.23	0.34	1
0.21	0.34	0.48	0.5	0.51	0.28	0.39	1
0.2	0.44	0.48	0.5	0.46	0.51	0.57	1
0.42	0.4	0.48	0.5	0.56	0.18	0.3	1

3.3 Features

The Ecoli Dataset contains seven attributes of continuous type, which represent specific biological constituents found in each ecoli sample. The attribute values are as follows:

1. mcg: McGeoch's method for signal sequence recognition.

2. gvh: von Heijne's method for signal sequence recognition.
3. lip: von Heijne's Signal Peptidase II consensus sequence score.
4. chg: Presence of charge on N-terminus of predicted lipoproteins.
5. aac: score of discriminant analysis of the amino acid content of outer membrane and periplasmic proteins.
6. alm1: score of the ALOM membrane spanning region prediction program.
7. alm2: score of ALOM program after excluding putative cleavable signal regions from the sequence.

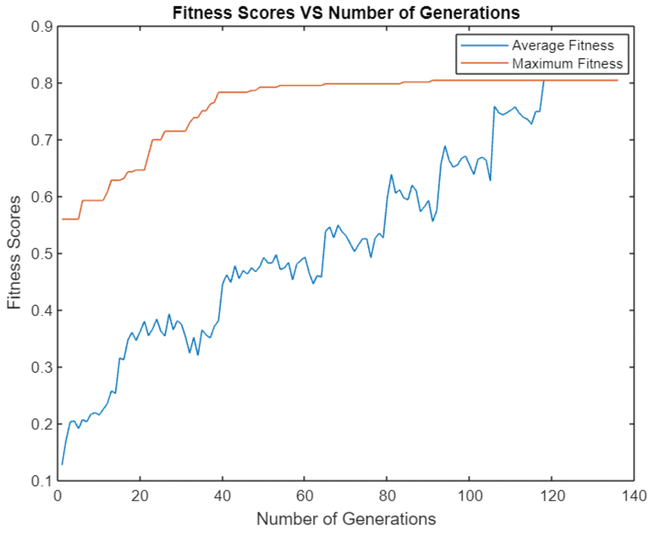
3.4 Classes

The dataset is divided into eight classes, each representing a different protein localization site. The class distribution is as follows:

Class 1: cp (cytoplasm) - 143 instances
Class 2: im (inner membrane without signal sequence) - 77 instances
Class 3: pp (perisplasm) - 52 instances
Class 4: imU (inner membrane, uncleavable signal sequence) - 35 instances
Class 5: om (outer membrane) - 20 instances
Class 6: omL (outer membrane lipoprotein) - 5 instances
Class 7: imL (inner membrane lipoprotein) - 2 instances
Class 8: imS (inner membrane, cleavable signal sequence) - 2 instances

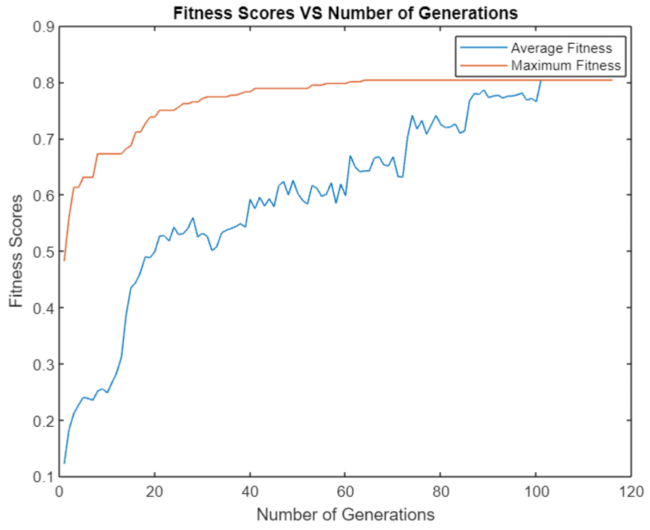
3.5 Results

	Generation =1		Generation = Last Generation			
SeedNumber	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for each Layer
496	0.12711	0.55952	0.80357	0.80357	**Avg**: 2 **Std**: 0	**Avg**: 7, 12 **Std**: 0, 0
966	0.12195	0.48214	0.80357	0.80357	**Avg**: 1 **Std**: 0	**Avg**: 10 **Std**: 0
570	0.11451	0.55655	0.70536	0.70536	**Avg**: 1 **Std**: 0	**Avg**: 8 **Std**: 0



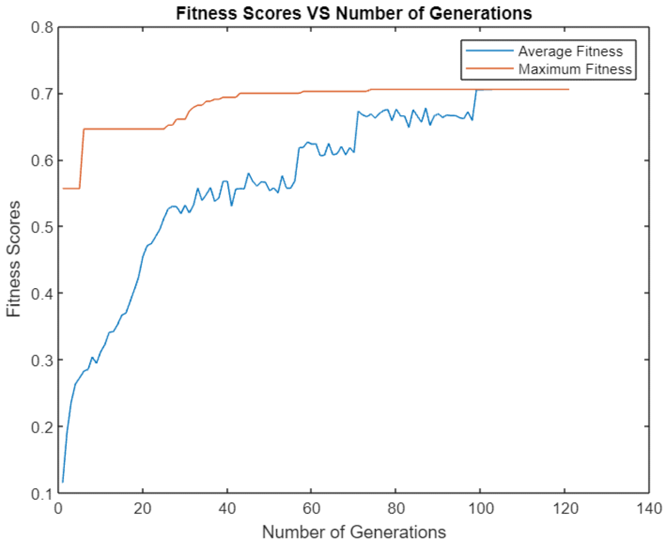
Graph 3.1 - Graph for Ecoli Dataset

Seed Number 496



Graph 3.2 - Graph for Ecoli Dataset

Seed Number 966



Graph 3.3 - Graph for Ecoli Dataset

Seed Number 570

3.6 Resources

The dataset is available online at

https://archive.ics.uci.edu/ml/datasets/ecoli