UNIVERSITI
MALAYA

ASSIGNMENT 1
MATLAB PROGRAMMING

WIX3001 SOFT COMPUTING

OCC 3

LECTURER: DR. LIEW WEI SHIUNG

NAME: WONG YAN JIAN

MATRIC NUMBER: U2102753/1

Table of Contents

1.0 Wine Recognition Dataset

1.1 Introduction

The Wine Recognition Dataset is a collection of data resulting from the chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The purpose of this dataset is to provide a basis for classification and correlation analysis of the wines based on their chemical composition.

1.2 Samples

The dataset contains a total of 178 samples, with each sample representing a wine from one of the three different cultivars.

Alcohol	Malic acid	Ash	Alcalinity of ash	Magnesium	Total phenols	Flavanoids	Nonflavanoid phenols
14.23	1.71	2.43	15.6	127	2.8	3.06	0.28
13.2	1.78	2.14	11.2	100	2.65	2.76	0.26
13.16	2.36	2.67	18.6	101	2.8	3.24	0.3
14.37	1.95	2.5	16.8	113	3.85	3.49	0.24
13.24	2.59	2.87	21	118	2.8	2.69	0.39
14.2	1.76	2.45	15.2	112	3.27	3.39	0.34
14.39	1.87	2.45	14.6	96	2.5	2.52	0.3
14.06	2.15	2.61	17.6	121	2.6	2.51	0.31
14.83	1.64	2.17	14	97	2.8	2.98	0.29
13.86	1.35	2.27	16	98	2.98	3.15	0.22
14.1	2.16	2.3	18	105	2.95	3.32	0.22

Proanthocyanins	Color intensity	Hue	OD280/OD315 of diluted wines	Proline	Class
2.29	5.64	1.04	3.92	1065	1
1.28	4.38	1.05	3.4	1050	1
2.81	5.68	1.03	3.17	1185	1
2.18	7.8	0.86	3.45	1480	1
1.82	4.32	1.04	2.93	735	1
1.97	6.75	1.05	2.85	1450	1
1.98	5.25	1.02	3.58	1290	1
1.25	5.05	1.06	3.58	1295	1
1.98	5.2	1.08	2.85	1045	1
1.85	7.22	1.01	3.55	1045	1
2.38	5.75	1.25	3.17	1510	1

1.3 Features

The dataset consists of 13 continuous attributes, which represent the quantities of specific chemical constituents found in each wine sample. The attributes included in the dataset are as follows:

1. Alcohol
2. Malic acid
3. Ash
4. Alkalinity of ash
5. Magnesium
6. Total phenols
7. Flavanoids
8. Nonflavanoid phenols
9. Proanthocyanins
10. Color intensity
11. Hue
12. OD280/OD315 of diluted wines
13. Proline

1.4 Classes

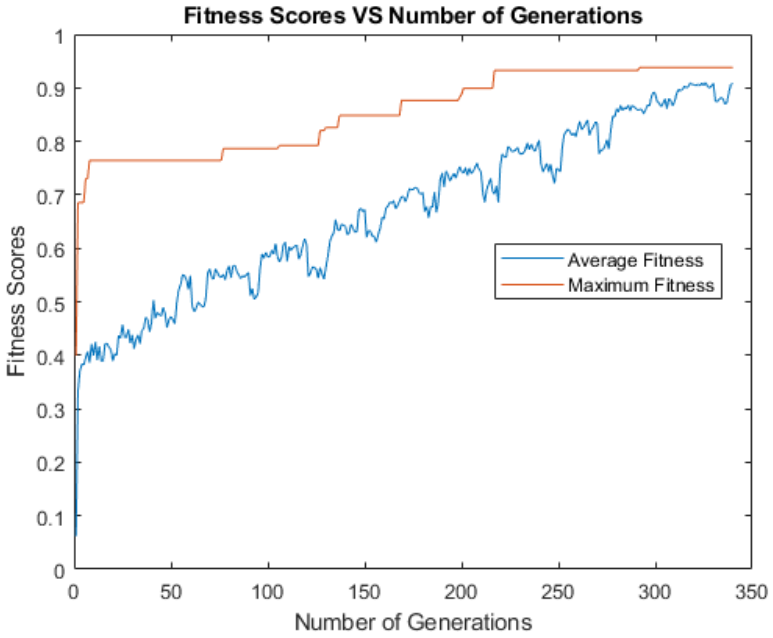
The dataset is divided into three classes, each representing a different cultivar of wine. The class distribution is as follows:

Class 1: 59 instances
Class 2: 71 instances
Class 3: 48 instances

1.5 Results

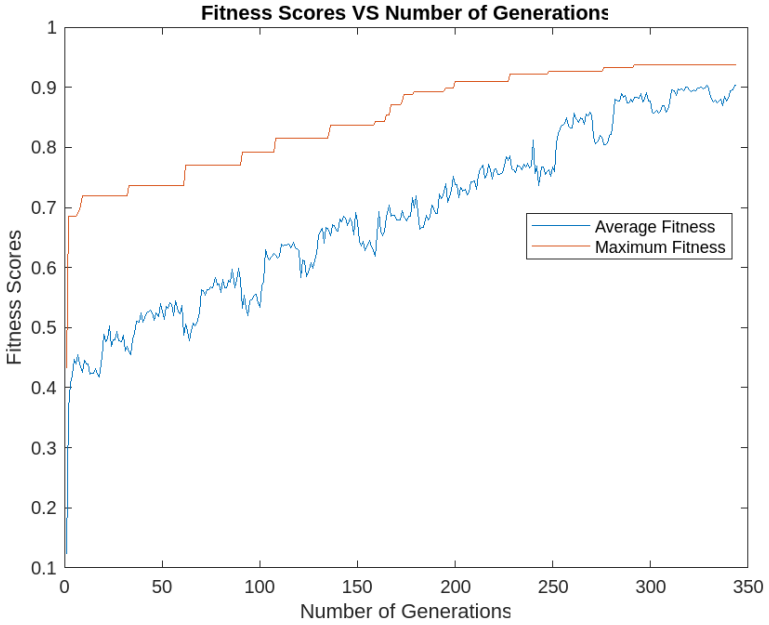
Belows are the results after running genetic algorithm on the dataset to search for the optimized solutions, including number of hidden layers, number of units per hidden layers, weights and biases.

	Generation = 1		Generation = Last Generation			
Seed Number	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for Each Layer
8320	0.06199	0.39889	0.90784	0.9382	**Avg**: 1 **Std**: 0	**Avg**: 4, 0, 0 ,0 ,0 **Std**: 0, 0, 0, 0, 0
9742	0.074466	0.39888	0.66739	0.76966	**Avg**: 2.71 **Std**: 1.17165	**Avg**: 38.01, 5.5, 23.63158, 10.03704, 2.80645 **Std**: 3.5059, 2.22141, 14.15177, 9.35861, 0.94585
6731	0.1239	0.43258	0.90315	0.9382	**Avg**: 1 **Std**: 0	**Avg**: 5, 0, 0, 0 **Std**: 0, 0, 0, 0



Graph 1.1 - Graph for Wine Recognition Dataset Seed Number 8320

Graph 1.2 - Graph for Wine Recognition Dataset Seed Number 9742



Graph 1.3 - Graph for Wine Recognition Dataset Seed Number 6731

1.6 Resources

The dataset is available online at https://archive-beta.ics.uci.edu/dataset/109/wine

2.0 Students' Dropout & Academic Success Dataset

2.1 Introduction

The Student's Dropout & Academic Success Dataset is a collection of data from several disjoint databases related to students enrolled in different undergraduate degrees, such as agronomy, design, education, nursing, journalism, management, social service, and technologies. The purpose of this dataset is to predict students' dropout and academic success using machine learning techniques, with the aim of reducing academic dropout and failure in higher education. It is supported by program SATDAP - Capacitação da Administração Pública under grant POCI-05-5762-FSE-000191, Portugal.

2.2 Samples

The dataset contains a total of 4424 samples, with each sample representing a student enrolled in an undergraduate degree program.

Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	Father's qualification	Mother's occupation	Father's occupation
1	17	5	171	1	1	122	1	19	12	5	9
1	15	1	9254	1	1	160	1	1	3	3	3
1	1	5	9070	1	1	122	1	37	37	9	9
1	17	2	9773	1	1	122	1	38	37	5	3
2	39	1	8014	0	1	100	1	37	38	9	9
2	39	1	9991	0	19	133.1	1	37	37	9	7
1	1	1	9500	1	1	142	1	19	38	7	10
1	18	4	9254	1	1	119	1	37	37	9	9
1	1	3	9238	1	1	137	62	1	1	9	9
1	1	1	9238	1	1	138	1	1	19	4	7

Admission grade	Displaced	Educational special needs	Debtor	Tuition fees up to date	Gender	Scholarship holder	Age at enrollment	International	Curricular units 1st sem (credited)	Curricular units 1st sem (enrolled)	Curricular units 1st sem (evaluations)	Curricular units 1st sem (approved)
127.3	1	0	0	1	1	0	20	0	0	0	0	0
142.5	1	0	0	0	1	0	19	0	0	6	6	6
124.8	1	0	0	0	1	0	19	0	0	6	0	0
119.6	1	0	0	1	0	0	20	0	0	6	8	6
141.5	0	0	0	1	0	0	45	0	0	6	9	5
114.8	0	0	1	1	1	0	50	0	0	5	10	5
128.4	1	0	0	1	0	1	18	0	0	7	9	7
113.1	1	0	0	0	1	0	22	0	0	5	5	0
129.3	0	0	0	1	0	1	21	1	0	6	8	6
123	1	0	1	0	0	0	18	0	0	6	9	5

Curricular units 1st sem (grade)	Curricular units 1st sem (without evaluations)	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)	Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)	Curricular units 2nd sem (without evaluations)	Unemployment rate	Inflation rate	GDP	Target
0	0	0	0	0	0	0	0	10.8	1.4	1.74	1
14	0	0	6	6	6	13.66666667	0	13.9	-0.3	0.79	2
0	0	0	6	0	0	0	0	10.8	1.4	1.74	1
13.42857143	0	0	6	10	5	12.4	0	9.4	-0.8	-3.12	2
12.33333333	0	0	6	6	6	13	0	13.9	-0.3	0.79	2
11.85714286	0	0	5	17	5	11.5	5	16.2	0.3	-0.92	2
13.3	0	0	8	8	8	14.345	0	15.5	2.8	-4.06	2
0	0	0	5	5	0	0	0	15.5	2.8	-4.06	1
13.875	0	0	6	7	6	14.14285714	0	16.2	0.3	-0.92	2
11.4	0	0	6	14	2	13.5	0	8.9	1.4	3.51	1

2.3 Features

The dataset consists of 36 attributes of different types, which include socio-economic factors, academic path, and academic performance metrics. The attributes included in the dataset are as follows:

1. Marital status
2. Application mode
3. Application order
4. Course
5. Daytime/evening attendance
6. Previous qualification
7. Previous qualification (grade)
8. Nationality
9. Mother's qualification
10. Father's qualification
11. Mother's occupation
12. Father's occupation
13. Admission grade
14. Displaced
15. Educational special needs
16. Debtor
17. Tuition fees up to date
18. Gender
19. Scholarship holder
20. Age at enrollment
21. International
22. Curricular units 1st sem (credited)
23. Curricular units 1st sem (enrolled)
24. Curricular units 1st sem (evaluations)
25. Curricular units 1st sem (approved)
26. Curricular units 1st sem (grade)
27. Curricular units 1st sem (without evaluations)
28. Curricular units 2nd sem (credited)
29. Curricular units 2nd sem (enrolled)
30. Curricular units 2nd sem (evaluations)
31. Curricular units 2nd sem (approved)
32. Curricular units 2nd sem (grade)
33. Curricular units 2nd sem (without evaluations)
34. Unemployment rate
35. Inflation rate
36. GDP

2.4 Classes

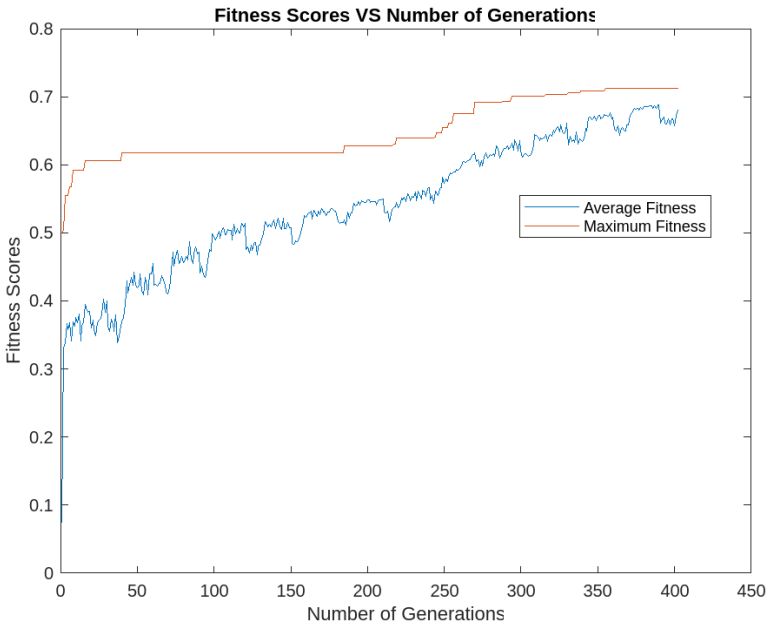
The dataset is a three category classification task, with the classes being dropout, enrolled and graduate, at the end of the normal duration of the course. The class distribution is as follows:

Class 1 - Dropout: 1421 instances
Class 2 - Graduate: 2209 instances
Class 3 - Enrolled: 794 instances

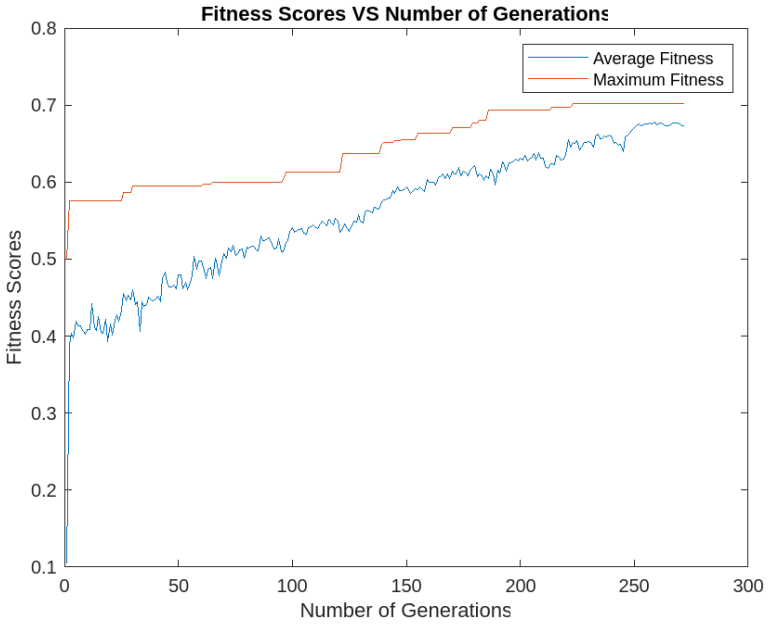
2.5 Results

Belows are the results after running genetic algorithm on the dataset to search for the optimized solutions, including number of hidden layers, number of units per hidden layers, weights and biases.

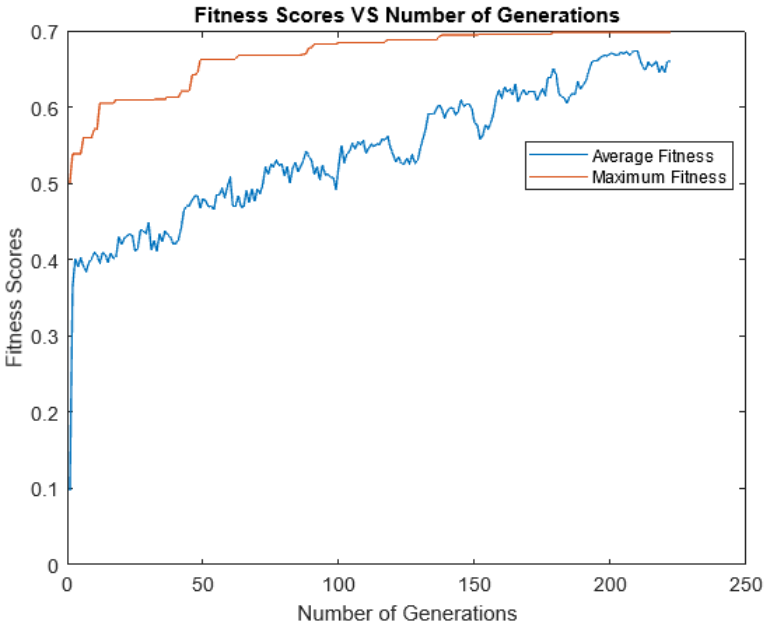
	Generation = 1		Generation = Last Generation			
Seed Number	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for Each Layer
9420	0.075506	0.49932	0.68069	0.7118	**Avg**: 1 **Std**: 0	**Avg**: 3, 0, 0, 0 **Std**: 0, 0, 0, 0
7234	0.10555	0.49977	0.67354	0.70253	**Avg**: 1 **Std**:0	**Avg**: 4, 0, 0, 0 **Std**: 0, 0, 0, 0
321	0.09805	0.50113	0.66001	0.69801	**Avg**: 1.035 **Std**: 0.27230	**Avg**: 2.005, 5, 3, 13 **Std**: 0.07071, 4.69042, 13, 0



Graph 2.1 - Graph for Students' Dropout & Academic Success Dataset Seed Number 9420



Graph 2.2 - Graph for Students' Dropout & Academic Success Dataset Seed Number 7234



Graph 2.3 - Graph for Students' Dropout & Academic Success Dataset Seed Number 321

2.6 Resources

The dataset is available online at https://archive-beta.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success

3.0 Car Evaluation Dataset

3.1 Introduction

The Car Evaluation Dataset is a dataset originally developed for the demonstration of DEX, an expert system for decision making. The dataset contains examples with the structural information removed and includes three intermediate concepts of a car, namely pricing, technical characteristics and comfort. The dataset is particularly useful for testing constructive induction and structure discovery methods due to its known underlying concept structure.

3.2 Samples

The dataset consists of 1728 samples, with each sample representing a car evaluated in terms of its intermediate concepts.

buying	maint	doors	persons	lug_boot	safety	class
4	4	2	2	1	1	1
4	4	2	2	1	2	1
4	4	2	2	1	3	1
4	4	2	2	2	1	1
4	4	2	2	2	2	1
4	4	2	2	2	3	1
4	4	2	2	3	1	1
4	4	2	2	3	2	1
4	4	2	2	3	3	1
4	4	2	4	1	1	1
4	4	2	4	1	2	1

3.3 Features

The Car Evaluation Dataset contains six attributes of categorical type, including buying, maint, doors, persons, lug_boot, and safety. The attribute values have been transformed to numerical for each attribute. The attribute values are as follows:

1. Buying (buying price): v-high - 4, high - 3, med - 2, low - 1
2. Maint (price of the maintenance): v-high - 4, high - 3, med - 2, low - 1
3. Doors (number of doors): 2, 3, 4, 5more - 5
4. Persons (capacity in terms of persons to carry): 2, 4, more - 6
5. Lug_boot (size of luggage boot): small - 1, med - 2, big - 3
6. Safety (estimated safety of the car): low - 1, med - 2, high - 3

3.4 Classes

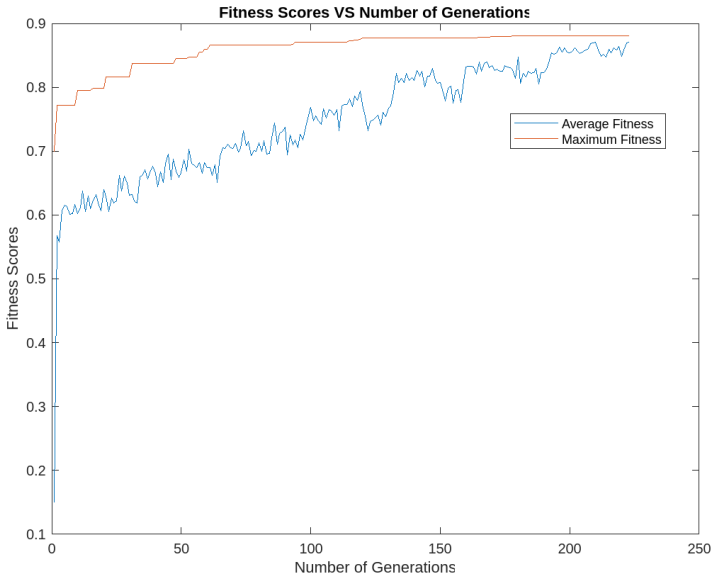
There are four classes in the dataset, namely unacc, acc, good, and v-good. Each value represents a different level of the car's acceptability. The class distribution is as follows:

Class 1 - unacc: 1210 instances
Class 2 - acc: 384 instances
Class 3 - good: 69 instances
Class 4 - vgood: 65 instances

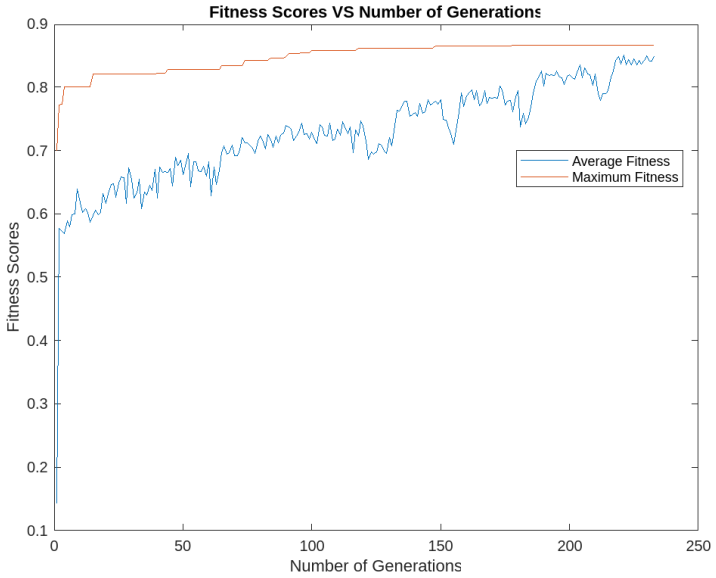
3.5 Result

Belows are the results after running genetic algorithm on the dataset to search for the optimized solutions, including number of hidden layers, number of units per hidden layers, weights and biases.

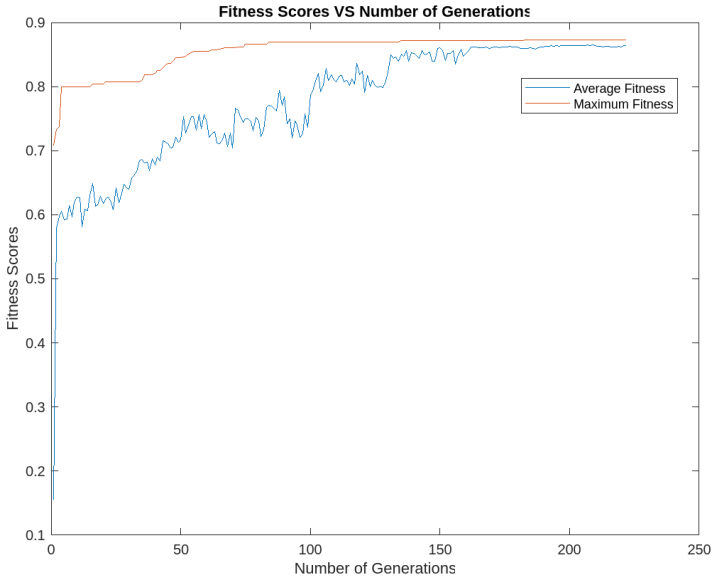
	Generation = 1		Generation = Last Generation			
Seed Number	AvgFitness	MaxFitness	AvgFitness	MaxFitness	Number of Layers	Number of Units for Each Layer
7652	0.15064	0.70023	0.87087	0.88079	**Avg**: 2.29 **Std**: 0.824194	**Avg**: 6, 2.005, 6.17391, 9.95652, 2.91667 **Std**: 0, 0.07071, 5.10483, 5.57982, 1.16450
6512	0.14315	0.70023	0.84961	0.86574	**Avg**: 1.225 **Std**: 0.79848	**Avg**: 7.02, 9.82353, 6.71429, 5.9, 4 **Std**: 0.14035, 5.17488, 4.76249, 5.30094, 2.82843
2343	0.15576	0.70891	0.8643	0.87326	**Avg**: 1 **Std**: 0	**Avg**: 5, 0, 0, 0 **Std**: 0, 0, 0, 0



Graph 3.1 - Graph for Car Evaluation Dataset Seed Number 7652



Graph 3.2 - Graph for Car Evaluation Dataset Seed Number 6512



Graph 3.3 - Graph for Car Evaluation Dataset Seed Number 2343

3.6 Resources

The dataset is available online at https://archive-beta.ics.uci.edu/dataset/19/car+evaluation