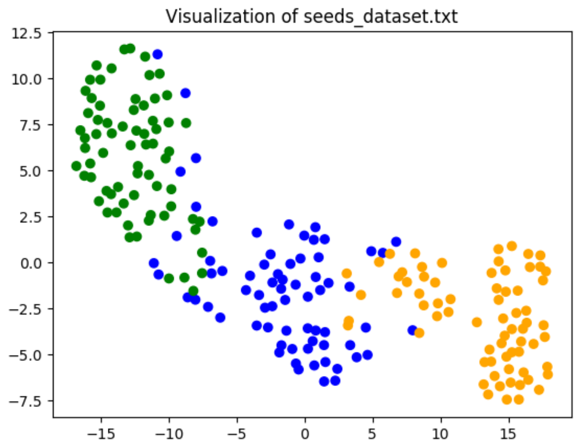
Assignment 4

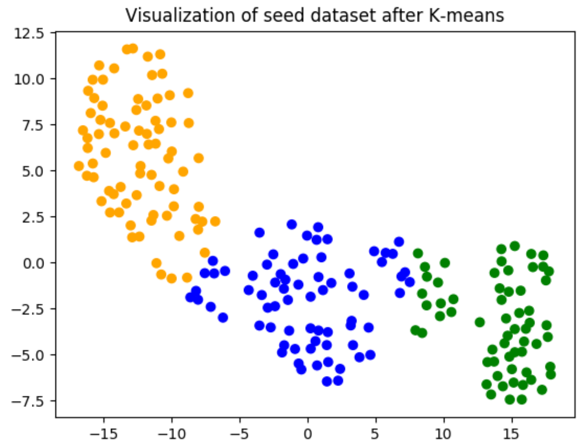
Machine Learning

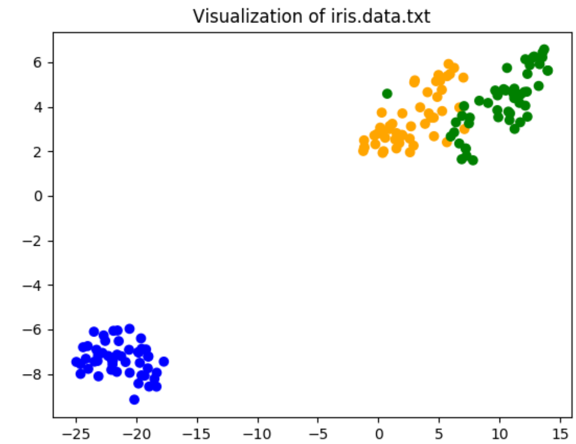
Submitted by: Himanshu Aggarwal
MT17015

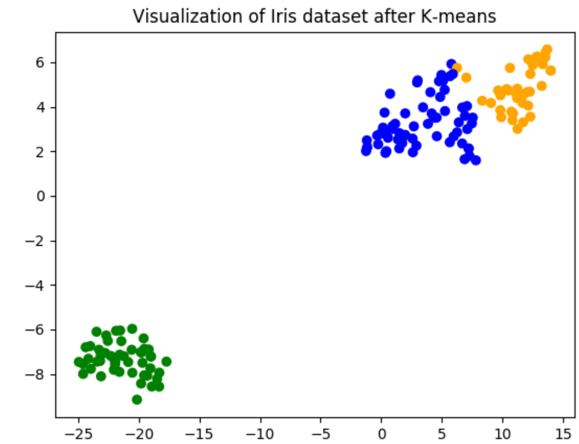
K-Means

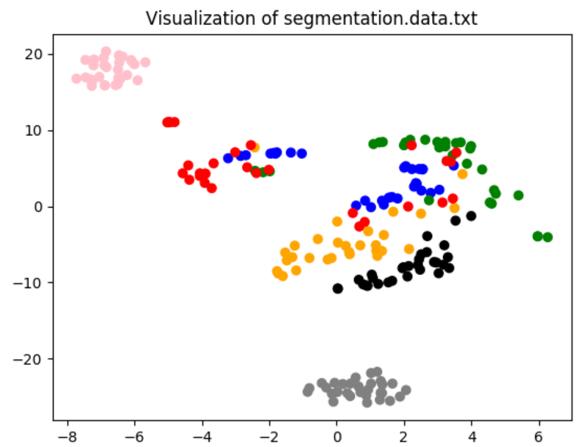
1. Visualisation and Qualitative metrics

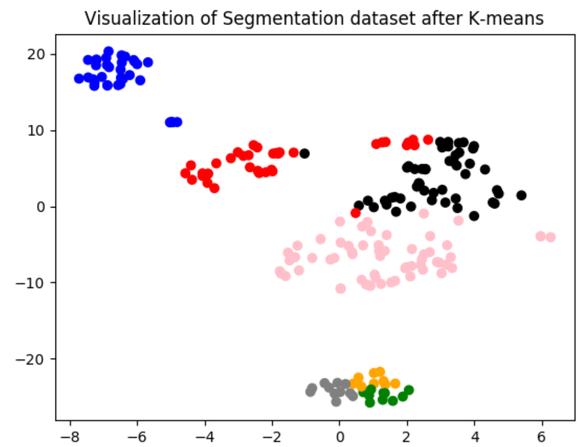


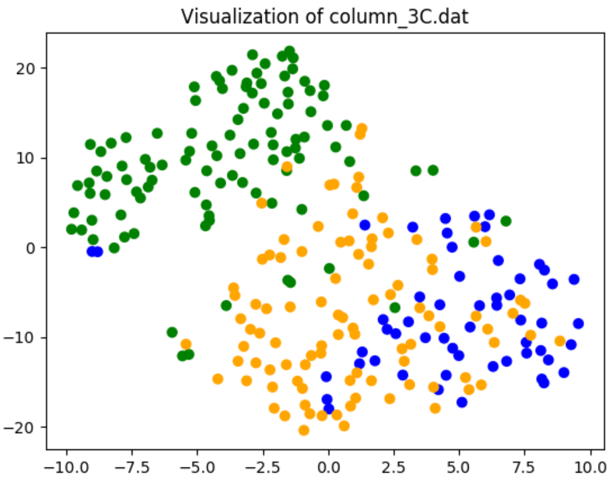


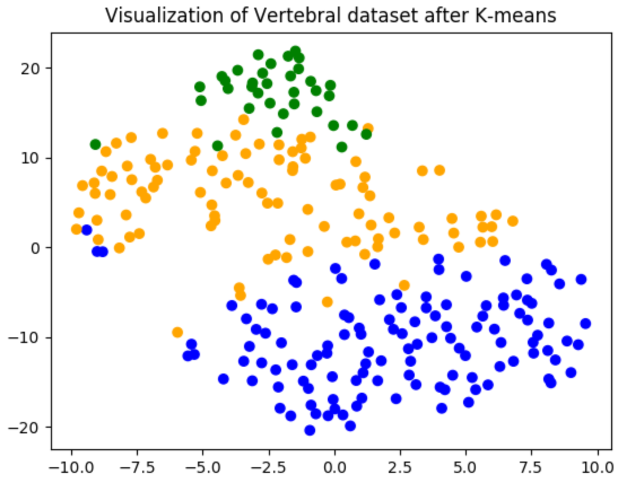


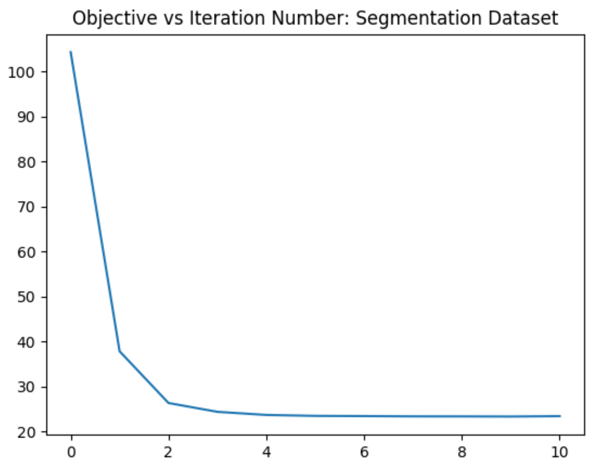


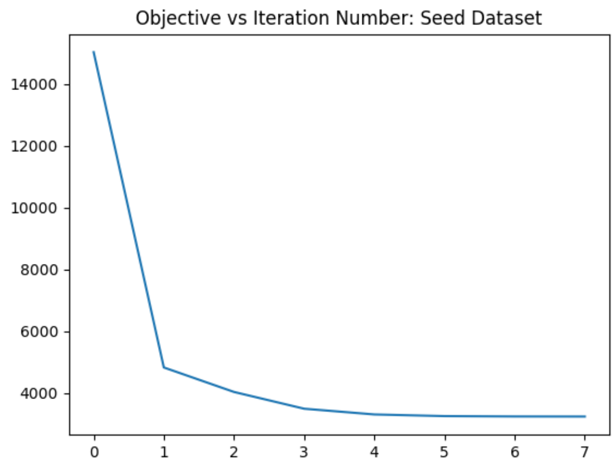


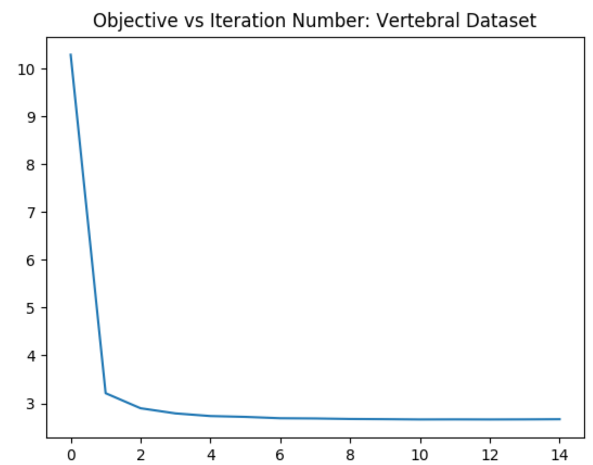


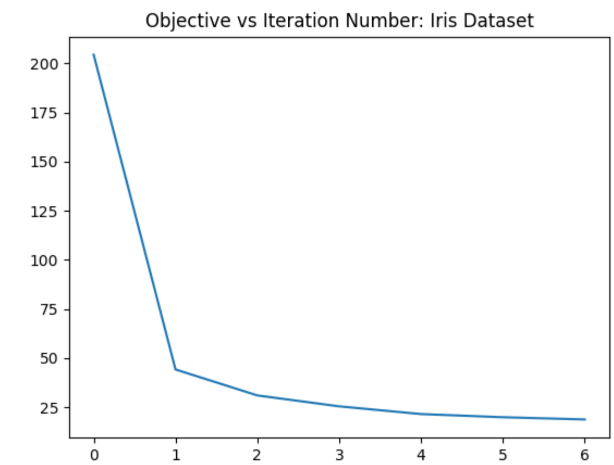

Visualization of column_3C.dat


Visualization of Vertebral dataset after K-means


Objective vs Iteration Number: Segmentation Dataset


Objective vs Iteration Number: Seed Dataset


Objective vs Iteration Number: Vertebral Dataset


Objective vs Iteration Number: Iris Dataset

Analysis:

- Clustering for **Seed** and **Iris** dataset was **near perfect**
- Clustering for **Segmentation** dataset was **fine**, not best. The data required some **preprocessing**. I have **removed useless features** from the data and **normalised** it, to have better clustering, but still there is noise.
- **Vertebral** dataset has a lot of **noise** in its data. Even after **normalisation** and **outlier removal**, as preprocessing techniques, the samples can be seen very noisy in the original scatter plot. Clusters formed for this dataset doesn't seem to be accurate due to large amount of noise present in the data.

2. Quantitative Metrics

Data	k = 2			k = true value			k = 12		
	ARI	NMI	AMI	ARI	NMI	AMI	ARI	NMI	AMI
Seed	3.57	3.50	3.48	3.57	3.50	3.48	3.564	3.52	3.496
Iris	2.697	3.249	3.044	2.695	3.254	3.045	2.989	3.409	3.253
Segmentation	1.475	2.719	1.88	1.414	2.656	1.843	1.471	2.591	1.825
Vertebral	1.163	1.516	1.461	1.122	1.472	1.399	1.167	1.513	1.45

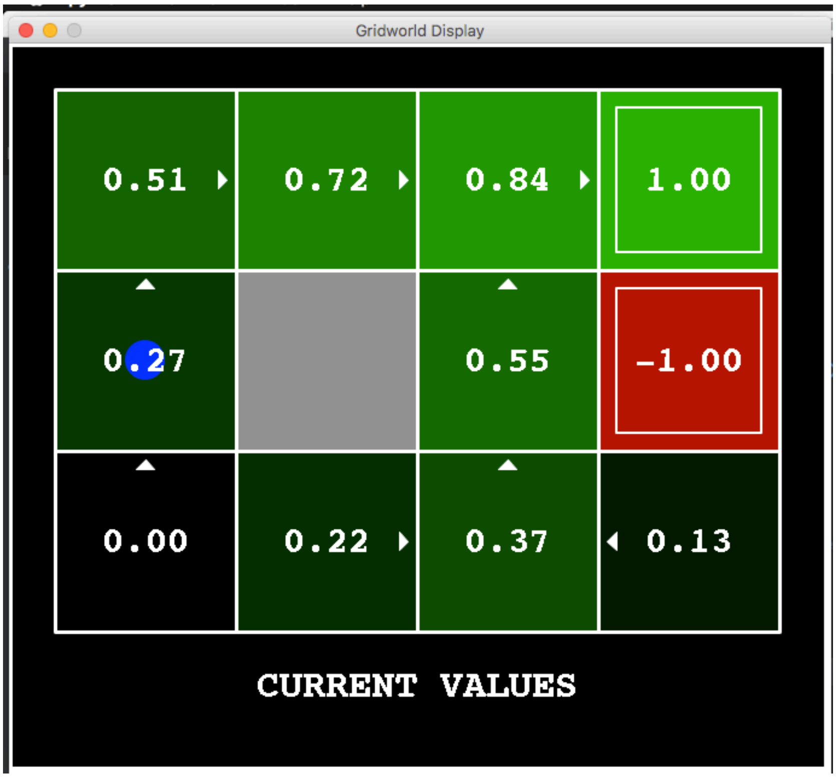
Analysis:

- From the above table it can be said that the clustering is good. Had there been negative or close to zero values, that would have represented bad clustering.
- ARI represents the similarity of the two assignments
- NMI and AMI measures the agreement of the two assignments.
- The first point applies for all the measures in the table.

Overall analysis says that the clustering done by K means for all the datasets is good, but for the Segmentation and Vertebral datasets, it can be improves upon. Better preprocessing techniques are required and K-means can also be applied feature wise and then the results can be improved.

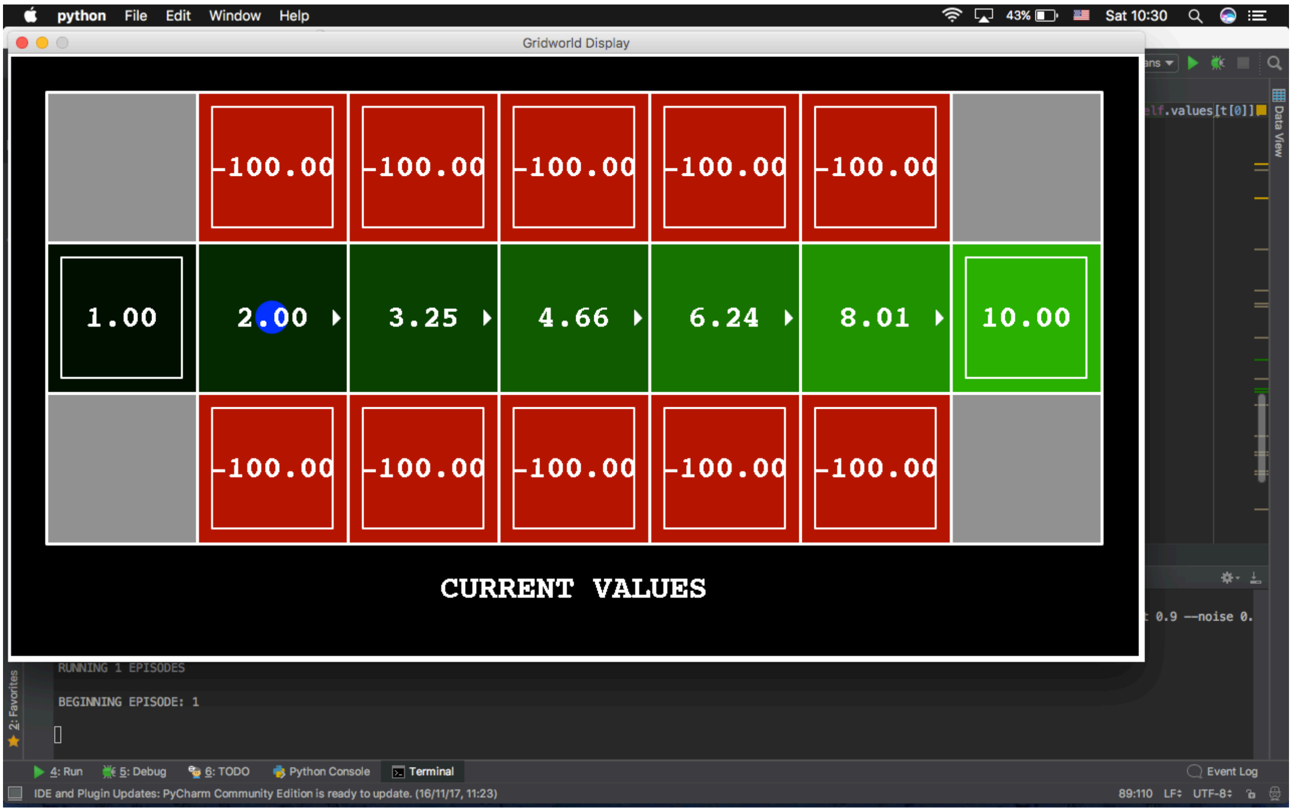
Reinforcement Learning

1. After writing Value Iteration Agent



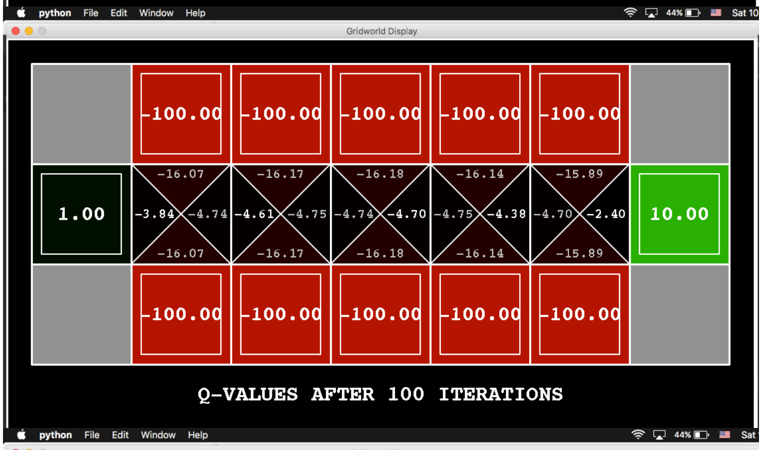
2. Bridge crossing

Correct Result with noise = 0.01 (or any value close to or equal to 0)

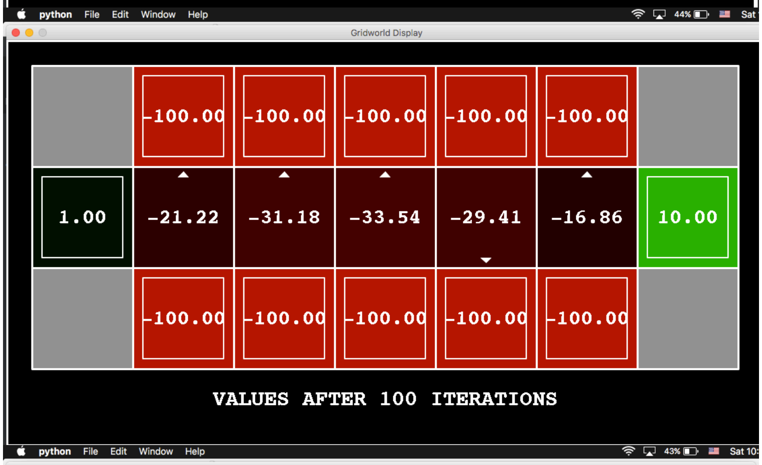


Analysis:
More noise makes agent act badly in the environment. Reducing the noise does the trick for us. Below are some hit-and-trials for this problem. We can see that changing discount do make the values of the states better than the default one, but high noise was still making the agent act badly.

Discount = 0.8



Discount = 0.2



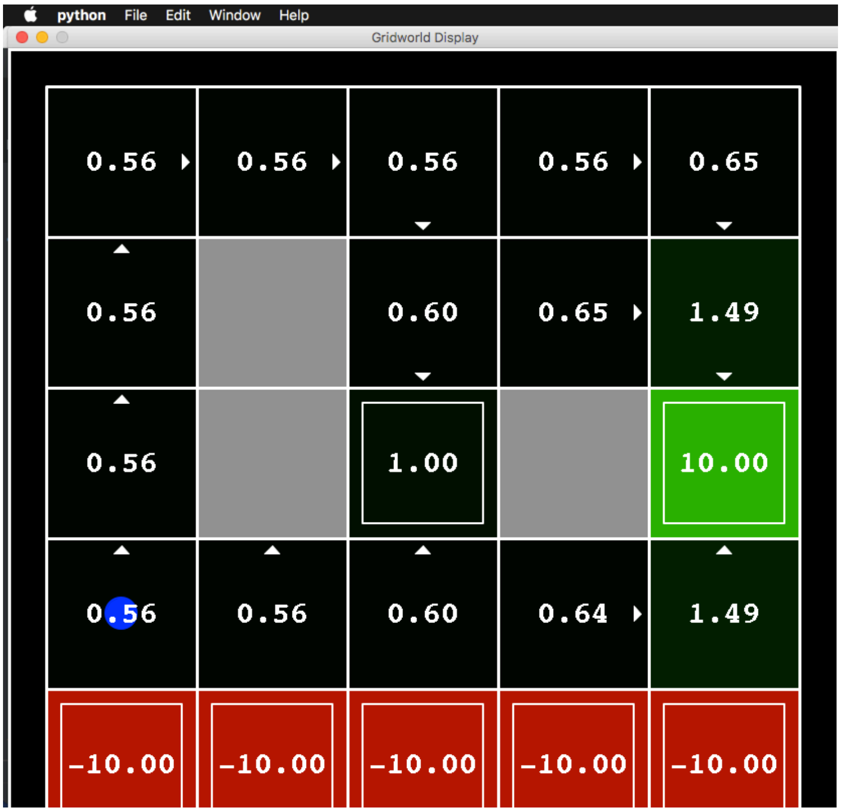
Noise = 0.9

Noise = 0

3. Discount Grid

i. Close exit risking cliff

ii. Close exit avoiding cliff



iii. Distant exit risking cliff

iv. Distant exit avoiding cliff

v. Never terminate

Analysis:

- **In 3.a, average discount and noise made the agent took the path we desire.**
- **In 3.b, reducing the discount and noise as compared to above point, and increasing living reward, made the agent follow desired path.**
- **In 3.c, removing the noise and low living reward makes the agent take optimal path to the best value that can be achieved.**
- **In 3.d, adding a little noise and increasing living reward, makes the agent take the longer path to the best value.**
- **In 3.e, making the noise max (= 1), makes the agent traverse the same path back and forth and never reach any desired state(goal).**

Above points highlights the importance of noise and living reward that can make the agent take different kinds of path(i.e make different kinds of policies).