CS 545 Project Proposal

Paresh Bhambhani and Benjamin Lickiss

November 6, 2015

1 Introduction

In Machine Learning there is a distinction between two types of algorithms: supervised and unsupervised. The supervised algorithms are methods that train a learning process using a confirmed result. This allows the training to have access to what is expected from the data. The unsupervised algorithms do not have access to this information while training, having to rely on the shape of the data alone.

In this project we propose a comparison of similar unsupervised and supervised techniques. We will be using classification methods on multi-class datasets, such as the digits set, to achieve a sufficiently high degree of complexity. For the first part of the project, the primary techniques being compared are k-means clustering with neural networks, with different initialization and implementation techniques used to vary the capabilities of each. For the second part, performance of k-means algorithm will be compared with that of particle swarm optimization based cluster algorithms.

These unsupervised algorithms have potential to be used in robotics applications for field learning, where the robot does not have access to human confirmation on what the data result should or should not be while training. The drawbacks of these may be a loss of accuracy or longer training time. Besides being used for unsupervised learning, PSO finds application in many facets of robotics including control of robotic manipulators and arms, motion planning and control, robot running, swarm navigation, swarm exploration and mapping and swarm obstacle avoidance. Ben and Paresh, both are PhD students working on their research in robotics. Hence working on this project, based on the concepts which have wide applicability in robotics, will help them explore and understand these topics better so that these may be employable in their research.

2 Refining K-Means Initial Points

The typical method of initializing the k-means clustering is to select the number of classes to be used, k, and then randomly create the center of each class, which is not always optimal for accuracy. To improve this, the initial points can be refined for better classification by finding multiple centers from subsets of the data[1]. The accuracy of the refined initial points, the randomized initial points, and a neural network with 1 and 2 hidden layers will be tested on the digits dataset.

To examine the effect scaling has on these methods, the Arcene dataset will be used as well. Both the full dataset and one with feature selection will be used to measure how well the methods scale to more complex data. All methods, k-means, k-means refined, and neural networks, will be tested.

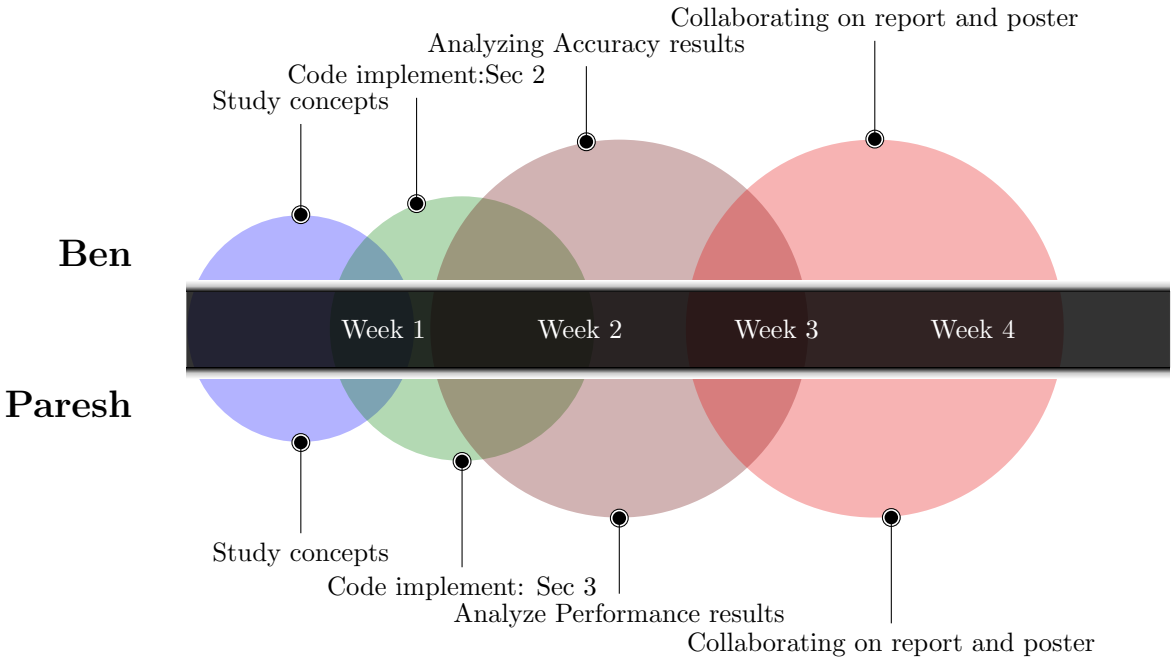
3 Data Clustering using Particle Swarm Optimization

PSO is stochastic search algorithm that maintains a population of particles, where each particle represents a potential solution, and aims at finding the particle position that results in the best evaluation of a given fitness function. In employing PSO for clustering, each particle represents cluster centroid vectors. A swarm of particles represents a number of candidate clusterings for the current data vectors. [3] [2]

Performance of PSO based cluster algorithm will be compared to the performance of K-Means algorithm on Wine dataset and digits dataset.

4 Proposed Timeline

The project is proposed to be completed in a month's time. A proposed timeline of the key steps is furnished below.



5 Team Members

Benjamin Lickiss: Benjamin is studying Robotics as a doctoral student at CSU. Currently he is examining the fault tolerant design of walking robots. His responsibilities in this project include the refined k-means method comparison and collaborating with Paresh on the report and poster.

Paresh Bhambhani: Paresh is a PhD student in Electrical Engineering department at CSU. His areas of interest include Swarm robot flocking, aggregation and exploration. Implementation and analysis of Section 3 which deals with the comparative study of the PSO based algorithm with the K-Means method will be covered by him besides collaborating with Ben on the report and poster.

References

[1] P. S. Bradley and U. M. Fayyad. Refining Iintial Points for K-Means Clustering. *Proceedings of the 15 th International Conference on Machine Learning (ICML98)*, pages 91–99, 1998.

[2] S. Cohen and L. N. de Castro. Data clustering with particle swarms. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1792–1798. IEEE, 2006.

[3] D. Van der Merwe and A. P. Engelbrecht. Data clustering using particle swarm optimization. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, volume 1, pages 215–220. IEEE, 2003.