CIS 419 – Homework 5

Devin Stein
devstein@seas.upenn.edu

November 30, 2016

Problem 1

Data: 4, 1, 9, 12, 6, 10, 2, 3, 9
First Iteration: Cluster Centers (Randomly Chosen): 1, 6
Data Assignments:
Cluster 1: 1, 2, 3
Cluster 2: 4, 6, 12, 6, 10, 9

a) Second Iteration:

Cluster 1 Center: 1/3 * (1 + 2 + 3) = 1/3 *6 = 2
Cluster 2 Center: 1/6 * (4 + 9 + 12 + 6 + 10 + 9) = 1/6 * 50 = 8.33
New Data Assignments:
Cluster 1: 1, 2, 3, 4
Cluster 2: 9, 12, 6, 10, 9

Third Iteration:

Cluster 1 Center: 1/4 * (1 + 2 + 3 + 4) = 1/4 * 10 = 2.5
Cluster 2 Center: 1/5 * (9 + 12 + 6 + 10 + 9) = 1/5 * 46 = 9.2
New Data Assignments:
Cluster 1: 1, 2, 3, 4
Cluster 2: 9, 12, 6, 10, 9

b) Yes we do converge as the cluster assignments are the same for both iteration.

Problem 2

a) Variance decreases as the value of K increases as there will be less outliers in each data assignments because of more possible clusters. Implying that all data will be closer to their assigned centroid.

b) When k=n! The algorithm will only converge when every cluster is equal to a unique point in the dataset. Each point is closest to itself.

Problem 3

It does not! If the agent does not successfully exit the maze it learns nothing. The reward function must be able to incrementally reward and penalize the agent. One such idea would be to give a -1 penalty to the agent if it does not successfully exit the maze. Another idea would be to give small rewards to the agent for discovering new areas in the maze and penalize the robot for repeating the same route more than twice.