Note: The source codes in C++ (written by Xuanhui Wang), Perl (written by Alex Kosorukoff) and Ocaml (written by Ramses Morales) are provided for the reference.

Note: For the next submission of source codes, if you name your package as firstName_hwN, it will be appreciated!

Note: 5 extra credit will be given if the homework is all printout and 2-3 extra credit will be given if the handwriting is clear.

Problem 1 (40 Points)

Problem 1.a (20 points)
Square distortion value is 7.48
Centroids:

\[
\begin{array}{cccccc}
2.10 & 2.20 & 1.00 & 3.10 & -1.50 \\
3.20 & 0.60 & 2.13 & -1.80 & 0.87 \\
-0.09 & 0.23 & 1.13 & 4.69 & -2.40
\end{array}
\]

Cluster Membership:
- cluster1: 
- cluster2: 3 6 9
- cluster3: 0 1 2 4 5 7 8

Problem 1.b (20 points)
Square distortion value is 8.34
Centroids:

\[
\begin{array}{cccccc}
2.58 & -0.93 & 1.38 & 0.83 & -0.80 \\
-1.62 & 2.25 & 1.50 & 5.60 & -2.35 \\
1.70 & 2.40 & 1.20 & 3.20 & -2.30
\end{array}
\]

Cluster Membership:
- cluster1: 1 3 4 6 7 9
- cluster2: 0 2 5 8
- cluster3:
Problem 1.c (10 points)

The results will be same as Problem 1.b since the algorithm has already converged.

Remark: Since the initial centroids are fixed, the clustering results (distortion value, centroid value and membership) in Problem a, b and c should be unique if you implement the standard K-means clustering algorithm. Two points will be deducted if your output in a), b) or c) are different from the standard output if you implement standard clustering algorithm.

Problem 2 (60 points)

Since T is not fixed, the output varies. So all points are given if you submit reasonable source code and printout. Some of your peer’s source codes are put online for reference.