HW4

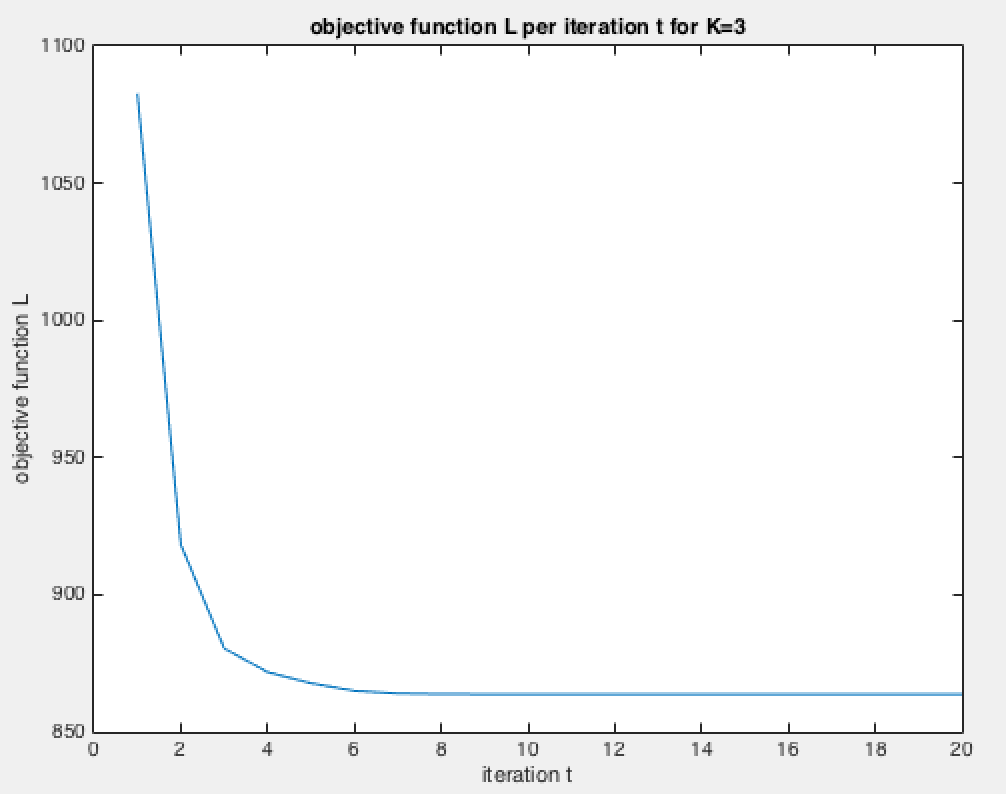
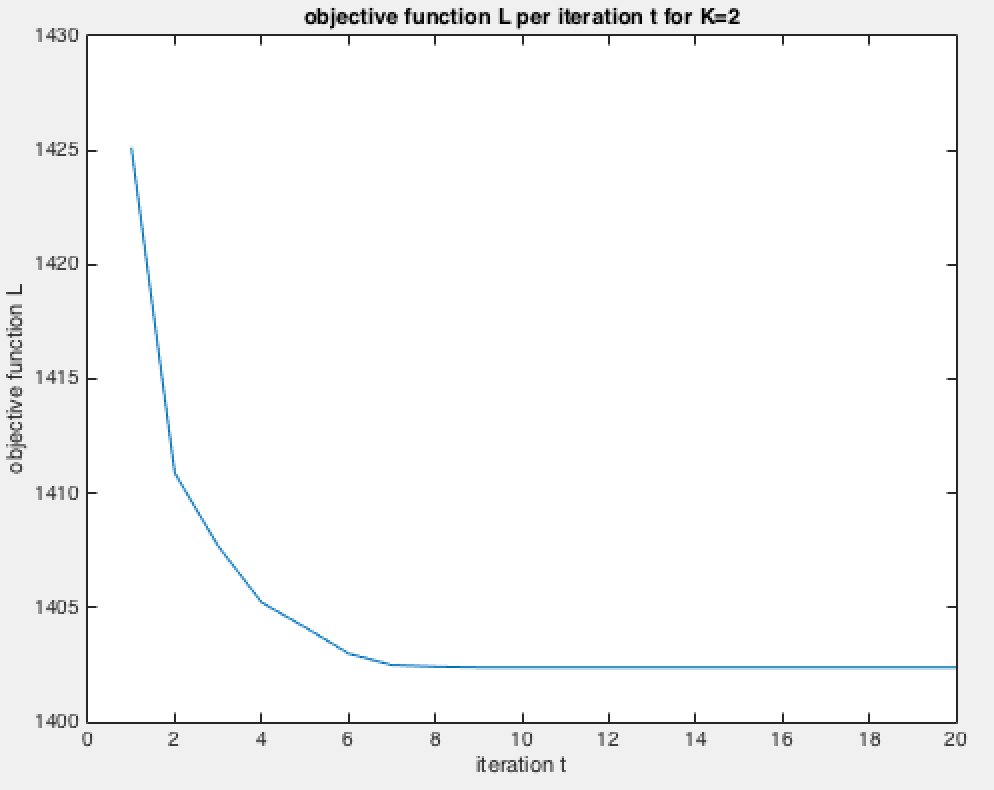
Shanqing Tan

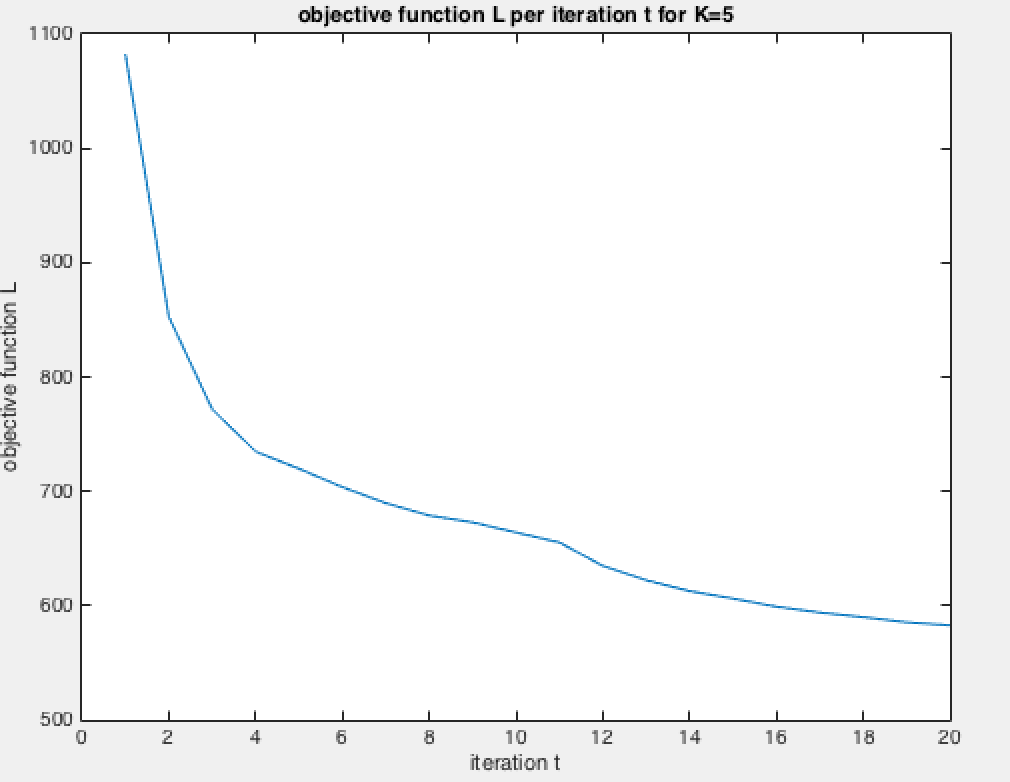
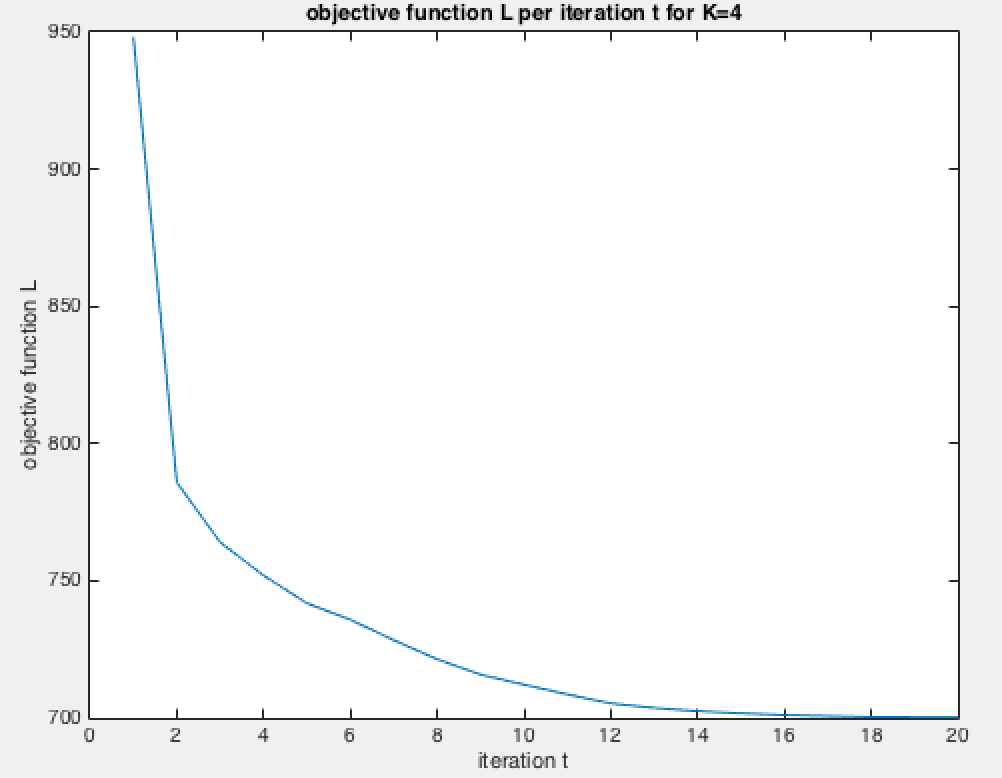
st2972

Problem1

1.For K = 2, 3, 4, 5, plot the value of the K-means objective function per iteration for 20 iterations.

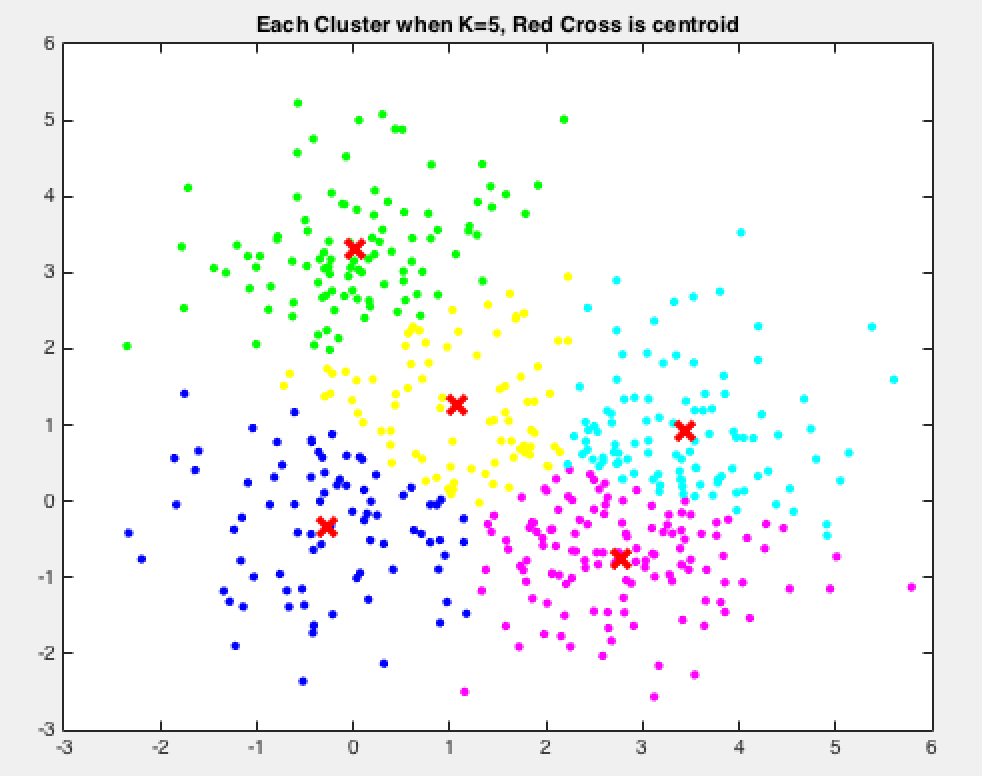
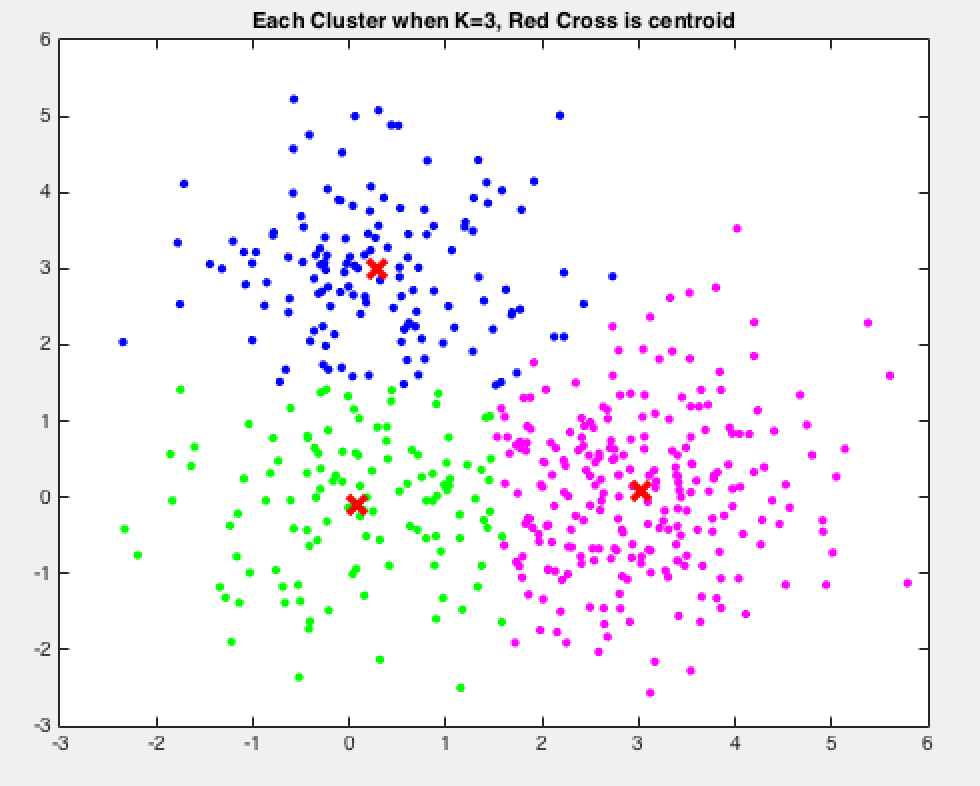
The pictures are as followed:





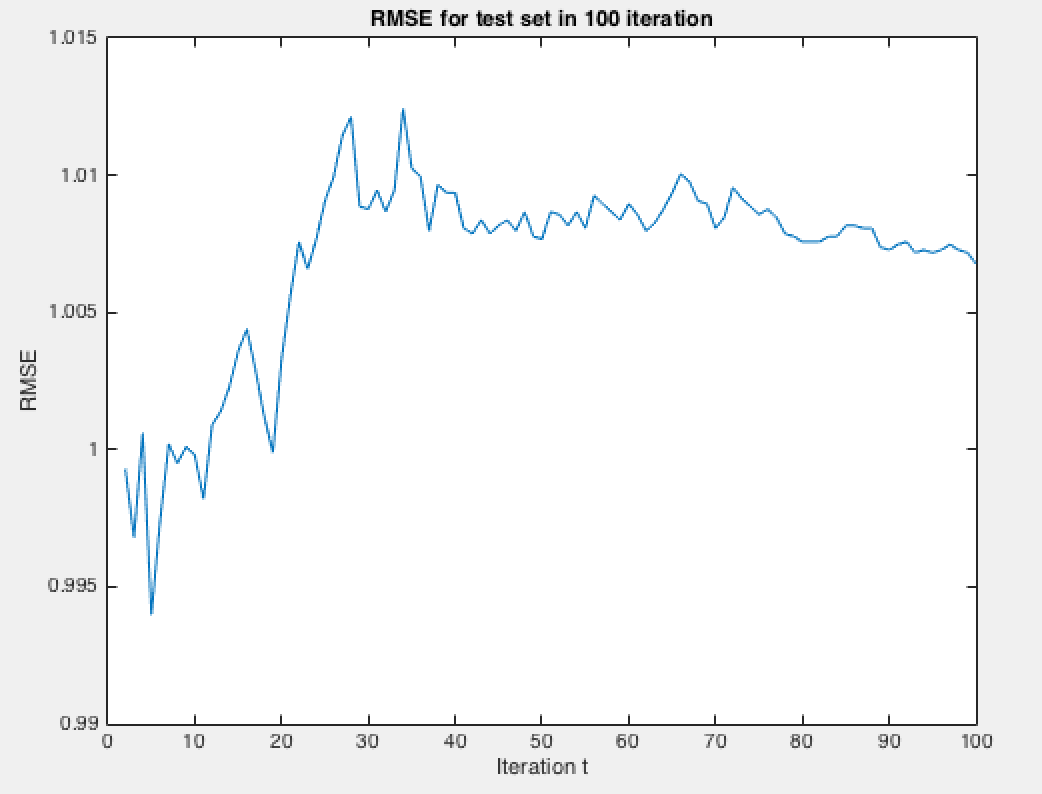
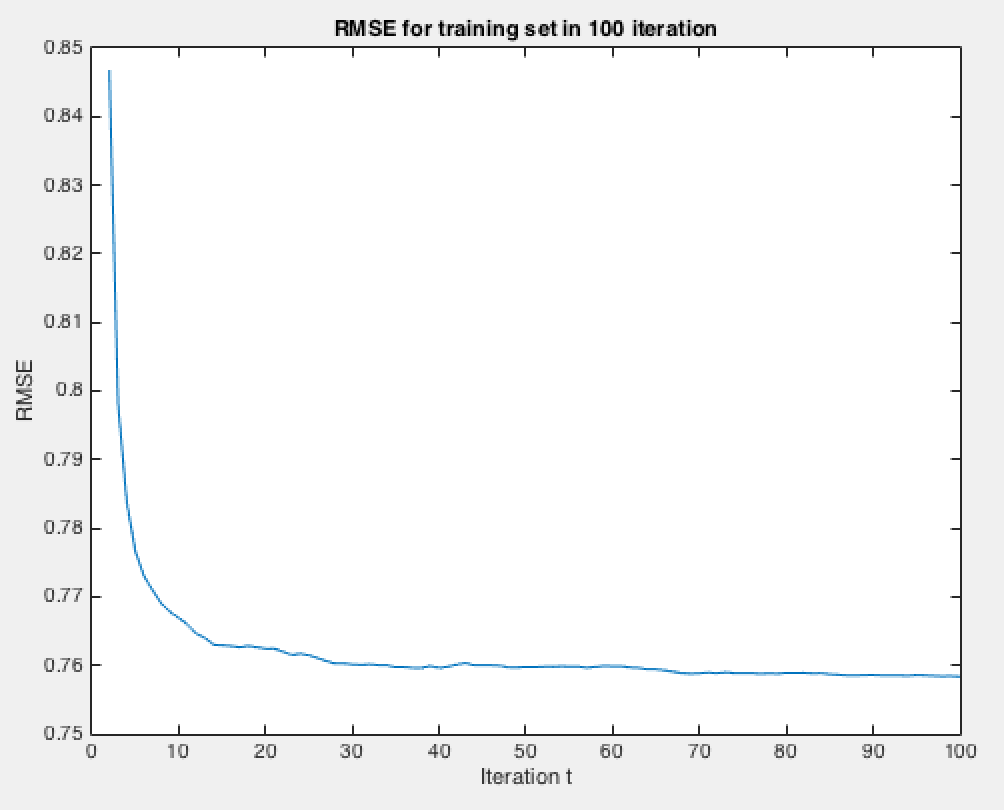
2. . For K = 3, 5, plot the 500 data points and indicate the cluster of each for the final iteration by marking it with a color or a symbol.

The centroids of the clusters are marked by red cross.

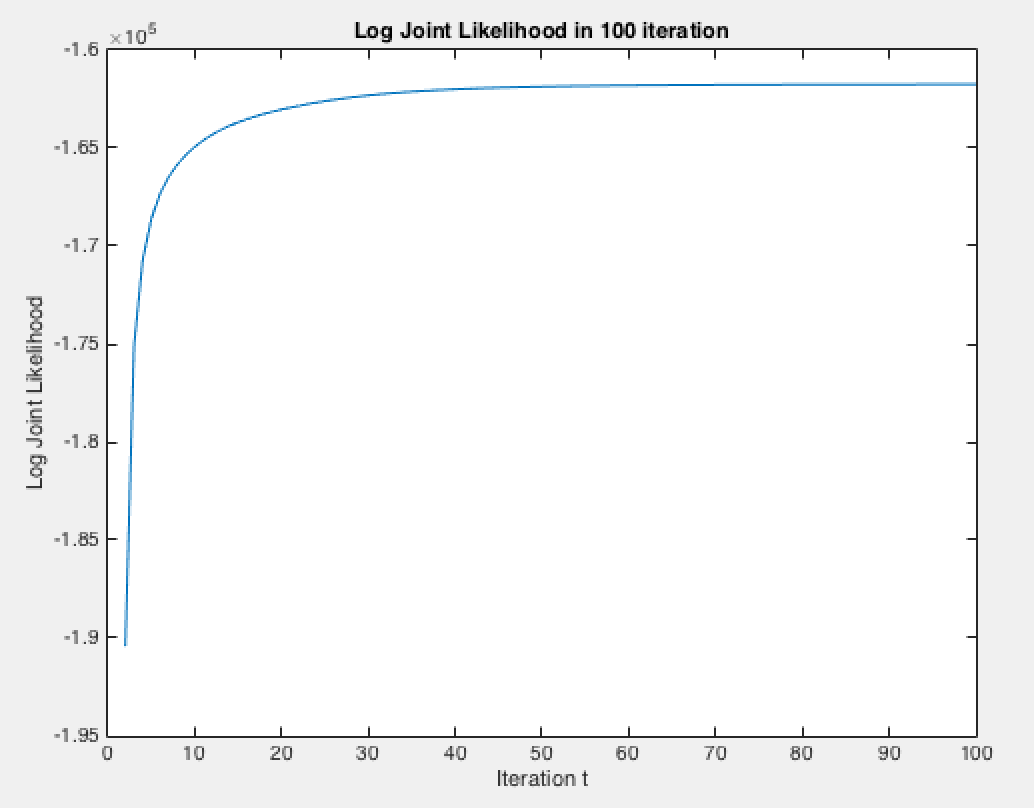


Problem2

1. . Plot the RMSE of your predictions on the held out test set provided for iteration number 2 to 100.



2. On a separate plot, show the log joint likelihood for iterations 2 to 100.



3. After 100 iterations, pick three reasonably well-known movies from the list provided and for each movie find the 5 closest movies according to Euclidean distance using their respective locations vj . List the query movie, the five nearest movies and their distances. A mapping from index to movie is provided with the data.

query movie: Star Wars (1977)

1st nearest movie: Empire Strikes Back, The (1980), distance=0.28458

2nd nearest movie: Return of the Jedi (1983), distance=0.51839

3rd nearest movie: Raiders of the Lost Ark (1981), distance=0.78797

4th nearest movie: Star Trek: The Wrath of Khan (1982), distance=1.0033

5th nearest movie: Aliens (1986), distance=1.0201

query movie: 101 Dalmatians (1996)

1st nearest movie: Indian in the Cupboard, The (1995), distance=0.477

2nd nearest movie: Santa Clause, The (1994), distance=0.51133

3rd nearest movie: Cool Runnings (1993), distance=0.55454

4th nearest movie: Father of the Bride Part II (1995), distance=0.56721

5th nearest movie: Hard Rain (1998), distance=0.5823

query movie: Air Force One (1997)

1st nearest movie: Courage Under Fire (1996), distance=0.62194

2nd nearest movie: Tombstone (1993), distance=0.63635

3rd nearest movie: Hunt for Red October, The (1990), distance=0.71576

4th nearest movie: Absolute Power (1997), distance=0.78829

5th nearest movie: Peacemaker, The (1997), distance=0.79571

4.

Centroid of Cluster 1 on u:

0.61211 0.26562 0.21326 -0.44437 0.37425 0.21599 -0.94866 1.3792 0.56699 0.66388

Number of Data in Cluster1=133

1st nearest movie: Godfather, The (1972), dot product=4.7044

2nd nearest movie: Shawshank Redemption, The (1994), dot product=4.5817

3rd nearest movie: Schindler's List (1993), dot product=4.5509

4th nearest movie: Pulp Fiction (1994), dot product=4.5475

5th nearest movie: Casablanca (1942), dot product=4.5233

6th nearest movie: Godfather: Part II, The (1974), dot product=4.5218

7th nearest movie: Boot, Das (1981), dot product=4.4904

8th nearest movie: One Flew Over the Cuckoo's Nest (1975), dot product=4.4889

9th nearest movie: Citizen Kane (1941), dot product=4.4673

10th nearest movie: Rear Window (1954), dot product=4.4577

Centroid of Cluster 2 on u:

0.11425 0.59983 0.32416 -0.62495 0.5149 0.47545 -0.66171 0.98308 0.46521 0.96345

Number of Data in Cluster2=120

1st nearest movie: Close Shave, A (1995), dot product=4.4931

2nd nearest movie: Schindler's List (1993), dot product=4.4509

3rd nearest movie: L.A. Confidential (1997), dot product=4.3598

4th nearest movie: Casablanca (1942), dot product=4.3595

5th nearest movie: Lone Star (1996), dot product=4.3465

6th nearest movie: Fargo (1996), dot product=4.338

7th nearest movie: Mrs. Brown (Her Majesty, Mrs. Brown) (1997), dot product=4.3374

8th nearest movie: Good Will Hunting (1997), dot product=4.3299

9th nearest movie: One Flew Over the Cuckoo's Nest (1975), dot product=4.3163

10th nearest movie: Full Monty, The (1997), dot product=4.3072

Centroid of Cluster 3 on u:

0.35671 -0.15572 -0.33038 -0.5864 0.54242 0.26694 -0.58522 1.0874 0.69456 0.96998

Number of Data in Cluster3=89

1st nearest movie: Shawshank Redemption, The (1994), dot product=4.5339

2nd nearest movie: Raiders of the Lost Ark (1981), dot product=4.483

3rd nearest movie: Schindler's List (1993), dot product=4.4534

4th nearest movie: Star Wars (1977), dot product=4.4364

5th nearest movie: Godfather, The (1972), dot product=4.4328

6th nearest movie: Bridge on the River Kwai, The (1957), dot product=4.4327

7th nearest movie: As Good As It Gets (1997), dot product=4.4242

8th nearest movie: Titanic (1997), dot product=4.4031

9th nearest movie: Silence of the Lambs, The (1991), dot product=4.3969

10th nearest movie: Great Escape, The (1963), dot product=4.3902

Centroid of Cluster 4 on u:

0.14341 -0.23826 -0.12393 -0.72223 0.28371 0.71418 -1.3812 1.1286 0.43261 0.59388

Number of Data in Cluster4=86

1st nearest movie: Wag the Dog (1997), dot product=4.7374

2nd nearest movie: Apt Pupil (1998), dot product=4.7084

3rd nearest movie: Titanic (1997), dot product=4.6679

4th nearest movie: Good Will Hunting (1997), dot product=4.5148

5th nearest movie: As Good As It Gets (1997), dot product=4.5035

6th nearest movie: Star Wars (1977), dot product=4.4373

7th nearest movie: L.A. Confidential (1997), dot product=4.4227

8th nearest movie: Schindler's List (1993), dot product=4.4172

9th nearest movie: Terminator 2: Judgment Day (1991), dot product=4.3666

10th nearest movie: Brassed Off (1996), dot product=4.3199

Centroid of Cluster 5 on u:

0.39054 0.60535 0.064694 -0.72033 0.46527 -0.24431 -0.65427 1.2281 0.47729 0.85392

Number of Data in Cluster5=80

1st nearest movie: Casablanca (1942), dot product=4.6226

2nd nearest movie: Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963), dot product=4.5932

3rd nearest movie: Citizen Kane (1941), dot product=4.5914

4th nearest movie: Lawrence of Arabia (1962), dot product=4.5803

5th nearest movie: Godfather, The (1972), dot product=4.5239

6th nearest movie: North by Northwest (1959), dot product=4.4485

7th nearest movie: Rear Window (1954), dot product=4.4431

8th nearest movie: 12 Angry Men (1957), dot product=4.4407

9th nearest movie: Third Man, The (1949), dot product=4.4374

10th nearest movie: Close Shave, A (1995), dot product=4.4321

5.

Centroid of Cluster 1 on v:

0.0044948 -0.39185 0.020565 -0.54303 0.038815 0.31181 -0.60772 0.65582 0.62314 0.36387

Number of Data in Cluster1=234

1st nearest movie: Last Action Hero (1993), distance=0.36137

2nd nearest movie: Pagemaster, The (1994), distance=0.37722

3rd nearest movie: Air Up There, The (1994), distance=0.38062

4th nearest movie: That Darn Cat! (1965), distance=0.41917

5th nearest movie: Swan Princess, The (1994), distance=0.43448

6th nearest movie: Moonlight and Valentino (1995), distance=0.43801

7th nearest movie: Life with Mikey (1993), distance=0.43806

8th nearest movie: Trial by Jury (1994), distance=0.45279

9th nearest movie: Stag (1997), distance=0.46296

10th nearest movie: Man of the House (1995), distance=0.46711

Centroid of Cluster 2 on v:

0.06259 -0.017923 0.039381 -0.085904 0.046759 0.091527 -0.23691 0.29597 0.17046 0.18731

Number of Data in Cluster2=213

1st nearest movie: Machine, The (1994), distance=0.18095

2nd nearest movie: Temptress Moon (Feng Yue) (1996), distance=0.19752

3rd nearest movie: Amityville 3-D (1983), distance=0.22642

4th nearest movie: Amityville 1992: It's About Time (1992), distance=0.23787

5th nearest movie: Amityville: A New Generation (1993), distance=0.23787

6th nearest movie: Venice/Venice (1992), distance=0.24407

7th nearest movie: Gordy (1995), distance=0.26275

8th nearest movie: Amityville Curse, The (1990), distance=0.26812

9th nearest movie: Amityville: Dollhouse (1996), distance=0.27161

10th nearest movie: Further Gesture, A (1996), distance=0.27429

Centroid of Cluster 3 on v:

0.25364 0.36329 0.14842 -0.60298 0.038758 0.079623 -0.64381 0.77687 0.337 0.54121

Number of Data in Cluster3=209

1st nearest movie: Nelly & Monsieur Arnaud (1995), distance=0.30026

2nd nearest movie: Love and Other Catastrophes (1996), distance=0.36933

3rd nearest movie: King of the Hill (1993), distance=0.37423

4th nearest movie: Angel and the Badman (1947), distance=0.38154

5th nearest movie: Feast of July (1995), distance=0.38938

6th nearest movie: Substance of Fire, The (1996), distance=0.3922

7th nearest movie: Wild Bill (1995), distance=0.41756

8th nearest movie: Purple Noon (1960), distance=0.42073

9th nearest movie: Man of the Year (1995), distance=0.4247

10th nearest movie: Object of My Affection, The (1998), distance=0.42488

Centroid of Cluster 4 on v:

0.51678 -0.072124 -0.088637 -0.34404 0.15355 0.21813 -0.60231 0.85011 0.26502 0.28974

Number of Data in Cluster4=179

1st nearest movie: Jerky Boys, The (1994), distance=0.31777

2nd nearest movie: Killer (Bulletproof Heart) (1994), distance=0.32087

3rd nearest movie: Underneath, The (1995), distance=0.3352

4th nearest movie: Bonheur, Le (1965), distance=0.36036

5th nearest movie: All Over Me (1997), distance=0.41502

6th nearest movie: Killer: A Journal of Murder (1995), distance=0.4164

7th nearest movie: Loaded (1994), distance=0.43048

8th nearest movie: Steal Big, Steal Little (1995), distance=0.43476

9th nearest movie: Substance of Fire, The (1996), distance=0.44003

10th nearest movie: House Party 3 (1994), distance=0.44766

Centroid of Cluster 5 on v:

0.081102 -0.38817 0.032928 -0.13162 0.4296 0.48854 -0.79671 0.80571 0.74738 0.75546

Number of Data in Cluster5=162

1st nearest movie: City of Angels (1998), distance=0.41199

2nd nearest movie: One Fine Day (1996), distance=0.4218

3rd nearest movie: Mrs. Doubtfire (1993), distance=0.44351

4th nearest movie: Losing Isaiah (1995), distance=0.48768

5th nearest movie: Abyss, The (1989), distance=0.49168

6th nearest movie: Prefontaine (1997), distance=0.50135

7th nearest movie: Mission: Impossible (1996), distance=0.51084

8th nearest movie: Outbreak (1995), distance=0.51624

9th nearest movie: It Could Happen to You (1994), distance=0.54185

10th nearest movie: Dangerous Minds (1995), distance=0.55931