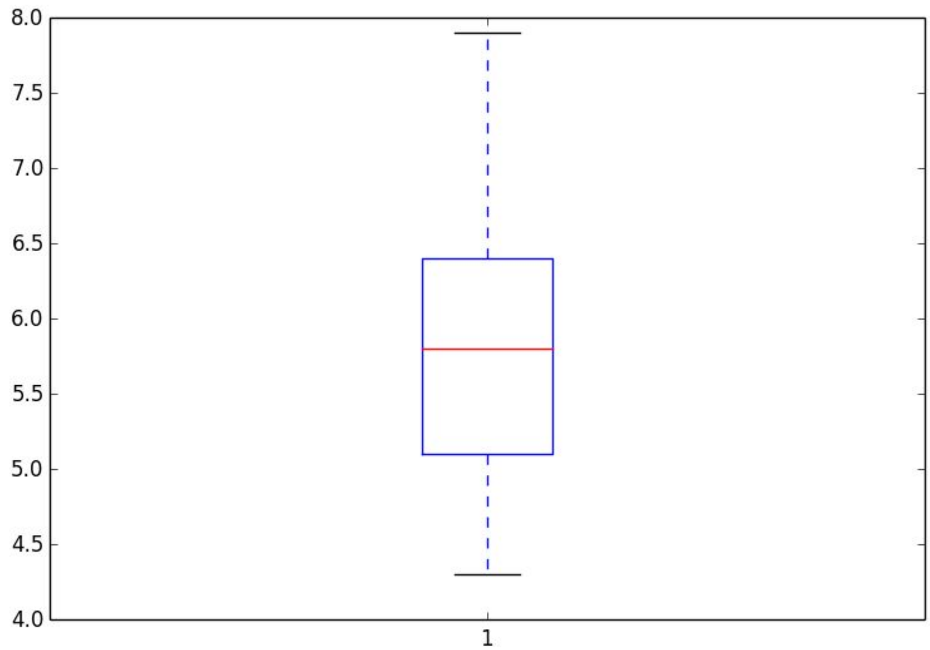
Shuai He
Programming HW1

Question 1

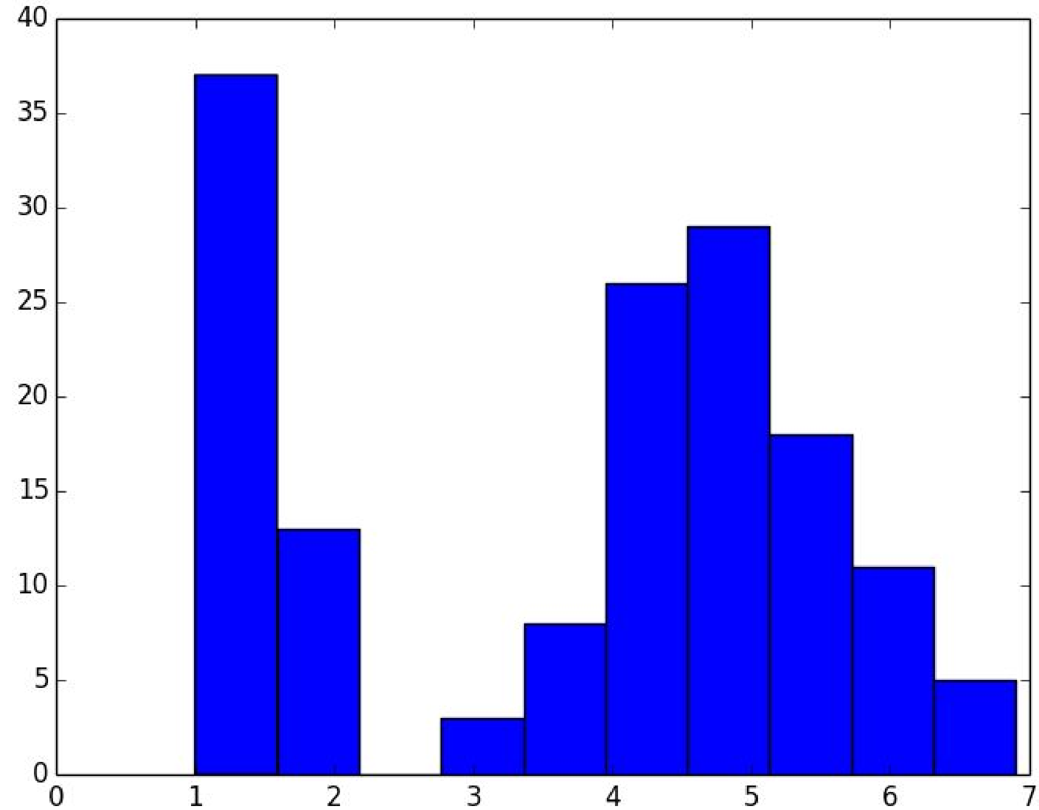
Feature	1	2	3	4
Max	2.4920192	3.09077525	1.78583195	1.71309869
Min	-1.87002413	-2.43394714	-1.56757623	-1.44954504

Question 2

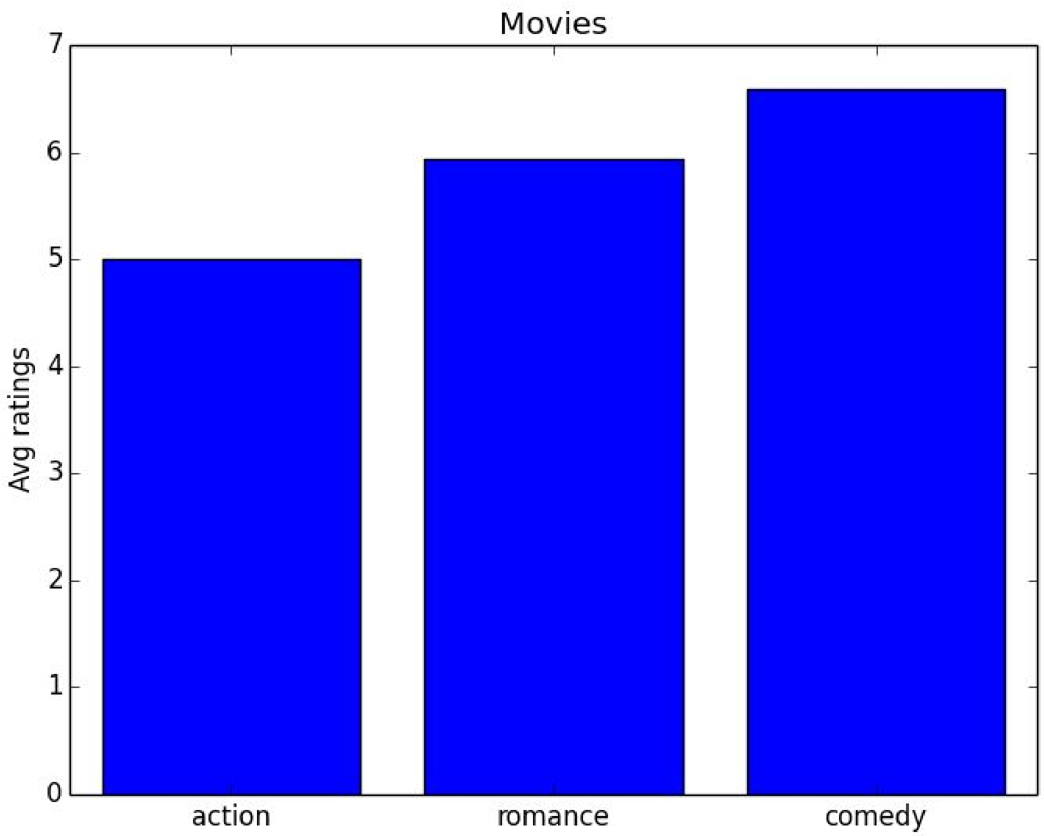
1) Boxplot (avg_website_1)



2) Histogram (avg_website_3)

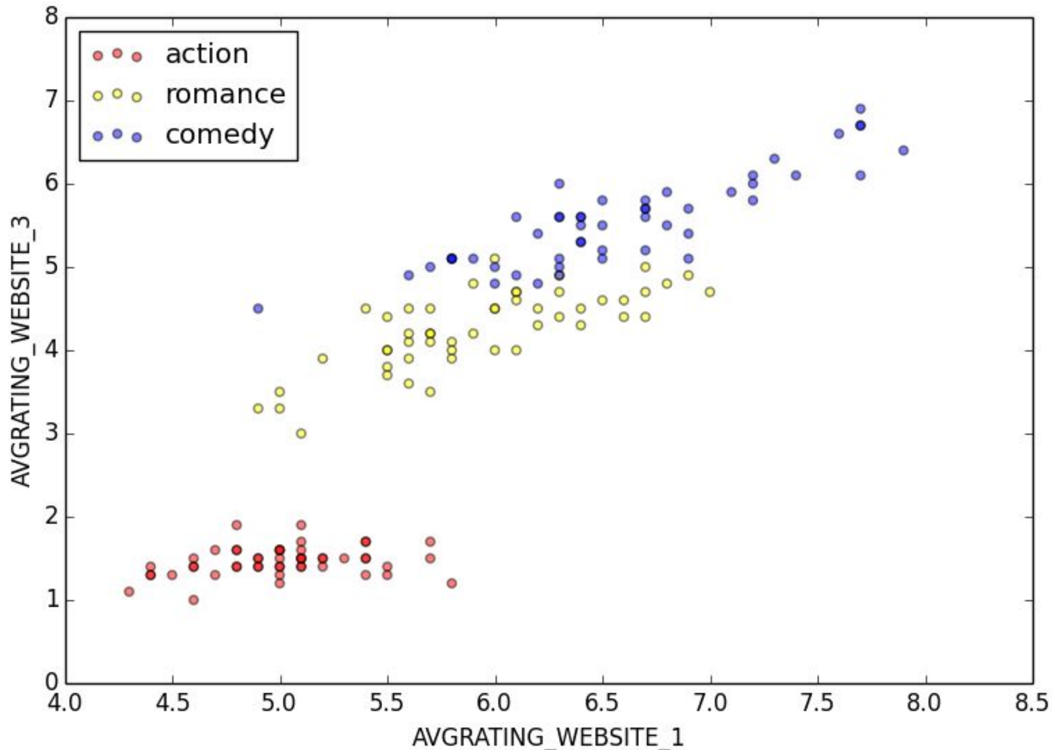


3) Bar Chart (avg_website_1)



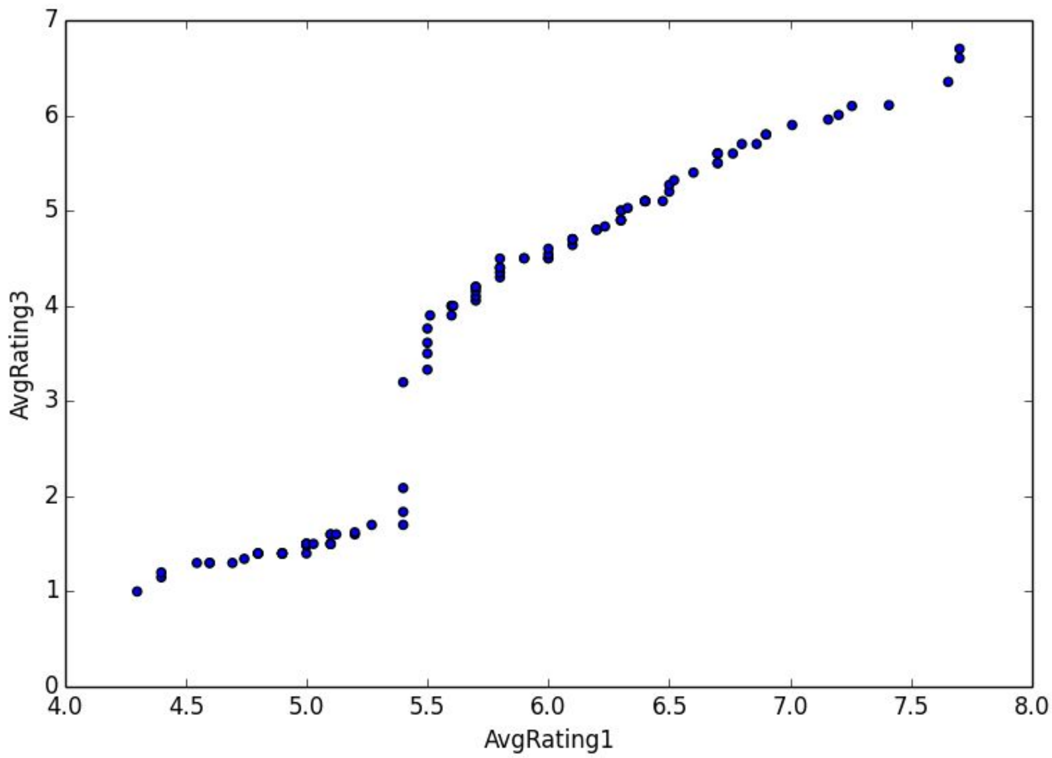
Movies
Avg ratings
0
1
2
3
4
5
6
7
action
romance
comedy

4)



action
romance
comedy
AVGRATING_WEBSITE_3
AVGRATING_WEBSITE_1
0
1
2
3
4
5
6
7
8
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5

5)



6) bins = [[1, 5),[5, 6),[6,7)]

KL_divergence(AVGRATING WEBSITE 1, AVGRATING WEBSITE 3) = 0.715706937172
KL_divergence(AVGRATING WEBSITE 3, AVGRATING WEBSITE 1) = 0.740031356772

Question 3

1)
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 2”) = -0.11756978
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 3”) = 0.87175378
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 4”) = 0.81794217
ρ(“AVGRATING WEBSITE 2”, “AVGRAT- ING WEBSITE 3”) = -0.4284401
ρ(“AVGRATING WEBSITE 2”, “AVGRAT- ING WEBSITE 4”) = -0.36543079
ρ(“AVGRATING WEBSITE 3”, “AVGRAT- ING WEBSITE 4”) = 0.96274602

2)
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 2”) = -0.11756978
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 3”) = 0.87175378
ρ(“AVGRATING WEBSITE 1”, “AVGRAT- ING WEBSITE 4”) = 0.81794217
ρ(“AVGRATING WEBSITE 2”, “AVGRAT- ING WEBSITE 3”) = -0.4284401
ρ(“AVGRATING WEBSITE 2”, “AVGRAT- ING WEBSITE 4”) = -0.36543079

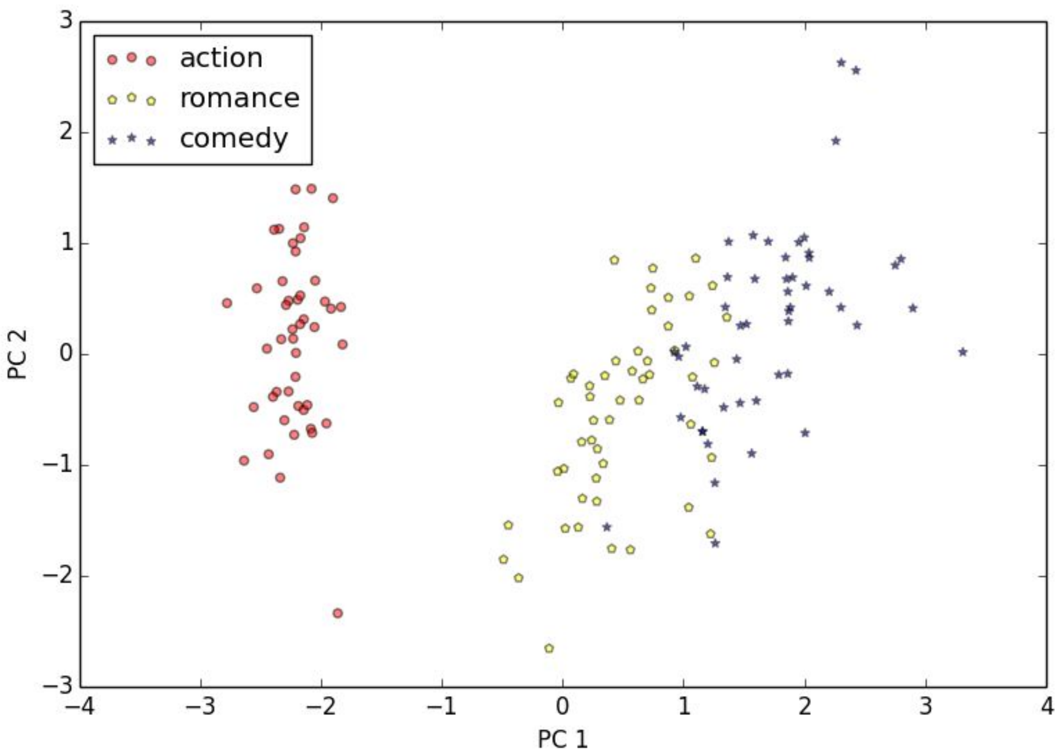
ρ(“AVGRATING WEBSITE 3”, “AVGRAT- ING WEBSITE 4”) = 0.96274602

3) Same.

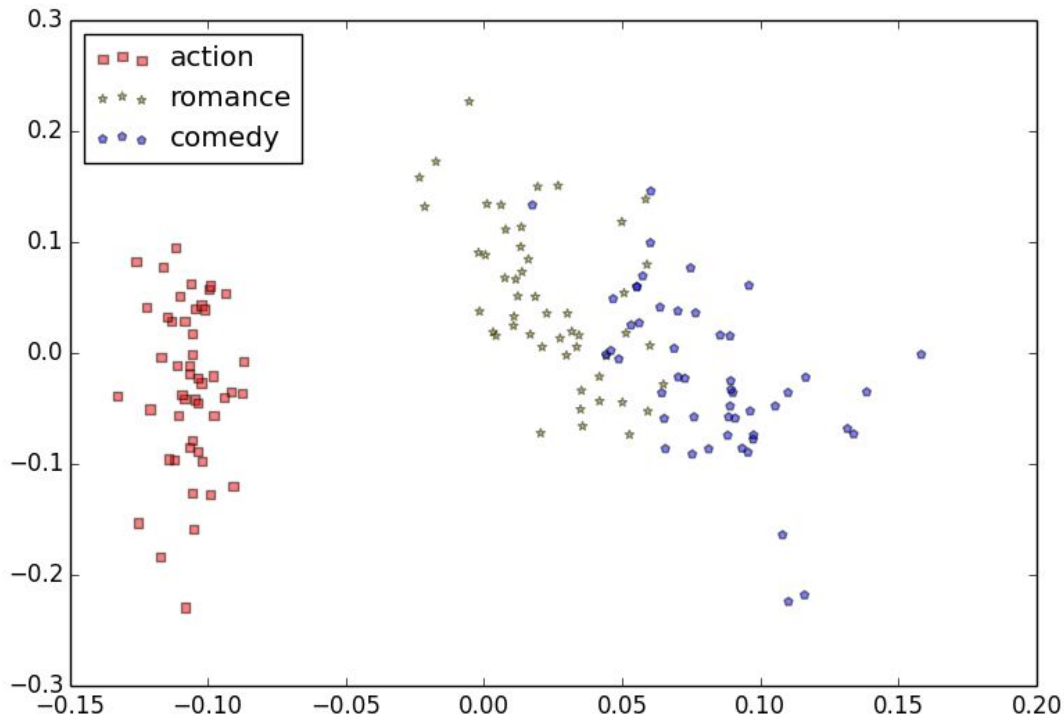
Because Corrcoef (Z-score) = Covariance(Z-score) = Corrcoef (Original)

Question 4

1) PCA:



2) SVD:



3)
Top-3 eigenvalues: [2.93779398 0.92025136 0.14793596]
Top-3 singular values: [20.92202913 11.70971619   4.69493963]

4) **It's the first singular vector, same as SVD.**
Because the propagation-based method approximates the SVD singular vectors. Based on the experiment result, as t gets bigger, u and v converge to the same as singular vectors.