



IRIS DATA CLUSTERING

Expectation maximization to Quantization maximization

G

```
Data generation : X  
Initialization : Y, small B, set A near 1  
HC = 0; Q=ceil(rand(N,1)*size(Y,1));
```

~HC

exit

*C:10

*F:21

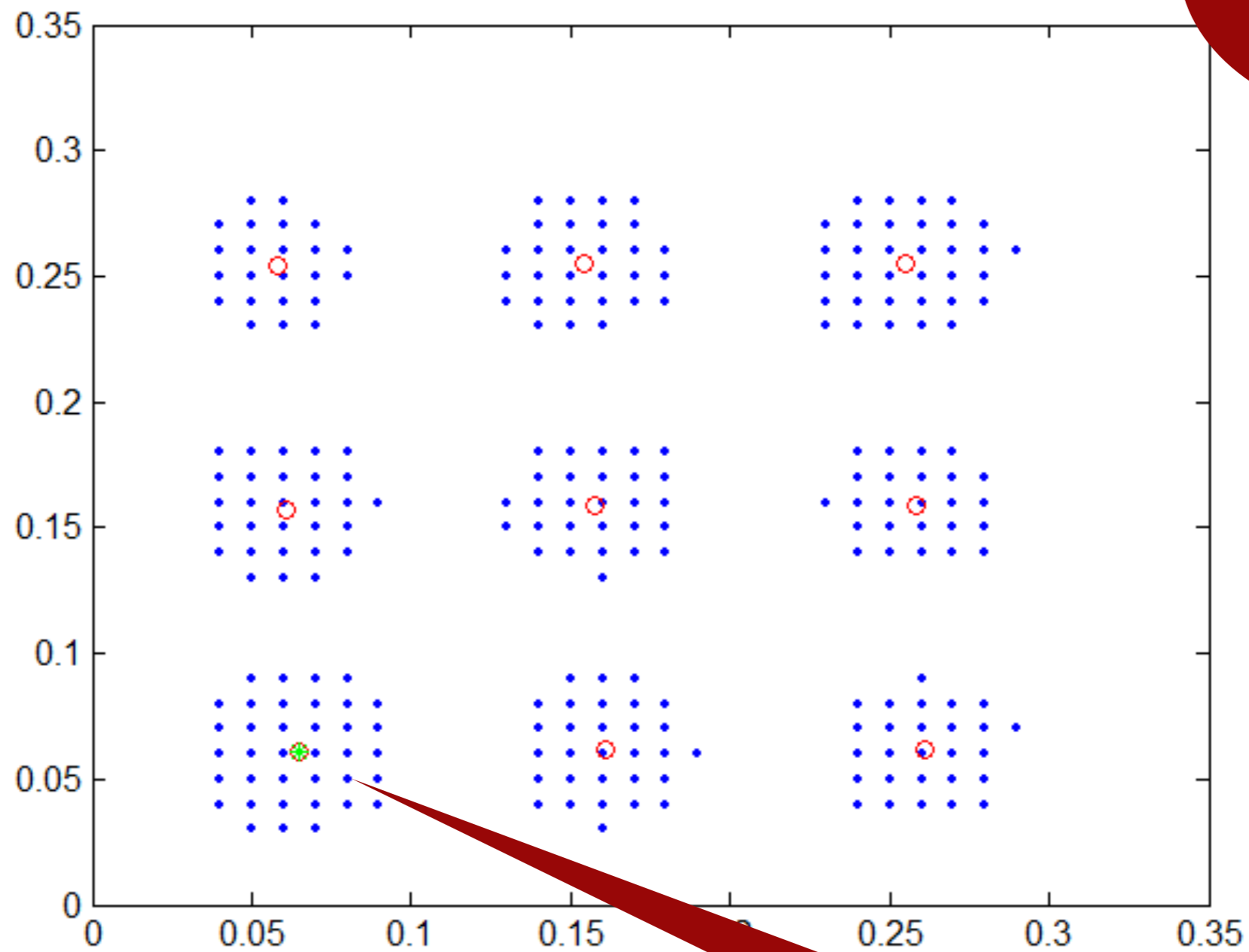
```
Calculate cross distances D  
Determine Q, stability and E  
Updating K centers Y  
if stability < 2/K Y=Y+rand(K,d)*0.02-0.01; end  
fprintf('B %f sta %f E %f\n',B stability,E);  
if stability > 0.98 HC=1; end  
B=B/A;
```

*E:17

$$\frac{1}{N} \sum_i \sum_k (g_k[i])^2$$

```
function R=demo_data9()
load data_9.mat;
K=10;
[C Q]=annealed_kmeans2(X,K);
save temp
[ss ind]=max(Q');
R=zeros(K,1);
for t=1:length(ind)
    i=ind(t);
    R(i)=R(i)+1;
end
plot(X(:,1),X(:,2),'.');hold on;plot(C(:,1),C(:,2),'ro');
ind=find(R==0);
plot(C(ind,1),C(ind,2),'g*');
```

sta 0.929389 E 0.000476



R: the number of data points in each center

ans =

25
29
34
27
28
23
30
29
37
0

A cluster that contains no data points

Overlapping memberships, Criterion,

E

```
U= exp(-B*D);  
S=sum(U,2);  
ind_zero=find(S<10^-10); %  
S(ind_zero)=10^-6; %  
  
Q=U./(S*ones(1,K));  
  
stability=sum(Q.^2,2)/(K-length(ind_zero)); %  
E=mean(sum(Q.*D.^2,2))
```

REVISE

150 IRIS DATA

Iris Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Famous database; from Fisher, 1936



Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	448407

Source:

Creator:

R.A. Fisher

- [LINK TO Website](#)



LINK

- . Number of Instances: 150 (50 in each of three classes)
- . Number of Attributes: 4 numeric, predictive attributes and the class
- . Attribute Information:
 1. sepal length in cm
 2. sepal width in cm
 3. petal length in cm
 4. petal width in cm
 5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica
- . Missing Attribute Values: None

Summary Statistics:

	Min	Max	Mean	SD	Class Correlation
sepal length:	4.3	7.9	5.84	0.83	0.7826
sepal width:	2.0	4.4	3.05	0.43	-0.4194
petal length:	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width:	0.1	2.5	1.20	0.76	0.9565 (high!)

- . Class Distribution: 33.3% for each of 3 classes.

5.1 3.5 1.4 0.2 0
4.9 3.0 1.4 0.2 0
4.7 3.2 1.3 0.2 0
4.6 3.1 1.5 0.2 0
5.0 3.6 1.4 0.2 0
5.4 3.9 1.7 0.4 0
4.6 3.4 1.4 0.3 0
5.0 3.4 1.5 0.2 0
4.4 2.9 1.4 0.2 0
4.9 3.1 1.5 0.1 0
5.4 3.7 1.5 0.2 0
4.8 3.4 1.6 0.2 0
4.8 3.0 1.4 0.1 0
4.3 3.0 1.1 0.1 0
5.8 4.0 1.2 0.2 0
5.7 4.4 1.5 0.4 0
5.4 3.9 1.3 0.4 0
5.1 3.5 1.4 0.3 0
5.7 3.8 1.7 0.3 0
5.1 3.8 1.5 0.3 0
5.4 3.4 1.7 0.2 0
5.1 3.7 1.5 0.4 0
4.6 3.6 1.0 0.2 0
5.1 3.3 1.7 0.5 0
4.8 3.4 1.9 0.2 0
5.0 3.0 1.6 0.2 0
5.0 3.4 1.6 0.4 0
5.2 3.5 1.5 0.2 0
5.2 3.4 1.4 0.2 0
4.7 3.2 1.6 0.2 0
4.8 3.1 1.6 0.2 0
5.4 3.4 1.5 0.4 0
5.2 4.1 1.5 0.1 0
5.5 4.2 1.4 0.2 0
4.9 3.1 1.5 0.1 0
5.0 3.2 1.2 0.2 0
5.5 3.5 1.3 0.2 0
4.9 3.1 1.5 0.1 0
4.4 3.0 1.3 0.2 0
5.1 3.4 1.5 0.2 0
5.0 3.5 1.3 0.3 0
6.3 2.8 5.1 1.5 2
6.1 2.6 5.6 1.4 2
7.7 3.0 6.1 2.3 2
6.3 3.4 5.6 2.4 2
6.4 3.1 5.5 1.8 2
6.0 3.0 4.8 1.8 2
6.9 3.1 5.4 2.1 2
6.7 3.1 5.6 2.4 2
6.9 3.1 5.1 2.3 2

5.8 2.7 5.1 1.9 2
6.8 3.2 5.9 2.3 2
6.7 3.3 5.7 2.5 2
6.7 3.0 5.2 2.3 2
6.3 2.5 5.0 1.9 2
6.5 3.0 5.2 2.0 2
6.2 3.4 5.4 2.3 2
5.9 3.0 5.1 1.8 2
7.1 3.0 5.9 2.1 2
6.3 2.9 5.6 1.8 2
6.5 3.0 5.8 2.2 2
7.6 3.0 6.6 2.1 2
4.9 2.5 4.5 1.7 2
7.3 2.9 6.3 1.8 2
6.7 2.5 5.8 1.8 2
7.2 3.6 6.1 2.5 2
6.5 3.2 5.1 2.0 2
6.4 2.7 5.3 1.9 2
6.8 3.0 5.5 2.1 2
5.7 2.5 5.0 2.0 2
5.8 2.8 5.1 2.4 2
6.4 3.2 5.3 2.3 2
6.5 3.0 5.5 1.8 2
7.7 3.8 6.7 2.2 2
7.7 2.6 6.9 2.3 2
6.0 2.2 5.0 1.5 2
6.9 3.2 5.7 2.3 2
5.6 2.8 4.9 2.0 2
7.7 2.8 6.7 2.0 2
6.3 2.7 4.9 1.8 2
6.7 3.3 5.7 2.1 2
7.2 3.2 6.0 1.8 2
6.2 2.8 4.8 1.8 2
6.1 3.0 4.9 1.8 2
6.4 2.8 5.6 2.1 2
7.2 3.0 5.8 1.6 2
7.4 2.8 6.1 1.9 2
7.9 3.8 6.4 2.0 2
6.4 2.8 5.6 2.2 2
4.5 2.3 1.3 0.3 0
4.4 3.2 1.3 0.2 0
5.0 3.5 1.6 0.6 0
5.1 3.8 1.9 0.4 0
4.8 3.0 1.4 0.3 0
5.1 3.8 1.6 0.2 0
4.6 3.2 1.4 0.2 0
5.3 3.7 1.5 0.2 0
5.0 3.3 1.4 0.2 0
7.0 3.2 4.7 1.4 1
6.4 3.2 4.5 1.5 1
6.9 3.1 4.9 1.5 1

5.5 2.3 4.0 1.3 1
6.5 2.8 4.6 1.5 1
5.7 2.8 4.5 1.3 1
6.3 3.3 4.7 1.6 1
4.9 2.4 3.3 1.0 1
6.6 2.9 4.6 1.3 1
5.2 2.7 3.9 1.4 1
5.0 2.0 3.5 1.0 1
5.9 3.0 4.2 1.5 1
6.0 2.2 4.0 1.0 1
6.1 2.9 4.7 1.4 1
5.6 2.9 3.6 1.3 1
6.7 3.1 4.4 1.4 1
5.6 3.0 4.5 1.5 1
5.8 2.7 4.1 1.0 1
6.2 2.2 4.5 1.5 1
5.6 2.5 3.9 1.1 1
5.9 3.2 4.8 1.8 1
6.1 2.8 4.0 1.3 1
6.3 2.5 4.9 1.5 1
6.1 2.8 4.7 1.2 1
6.4 2.9 4.3 1.3 1
6.6 3.0 4.4 1.4 1
6.8 2.8 4.8 1.4 1
6.7 3.0 5.0 1.7 1
6.0 2.9 4.5 1.5 1
5.7 2.6 3.5 1.0 1
5.5 2.4 3.8 1.1 1
5.5 2.4 3.7 1.0 1
5.8 2.7 3.9 1.2 1
6.0 2.7 5.1 1.6 1
5.4 3.0 4.5 1.5 1
6.0 3.4 4.5 1.6 1
6.7 3.1 4.7 1.5 1
6.3 2.3 4.4 1.3 1
5.6 3.0 4.1 1.3 1
5.5 2.5 4.0 1.3 1
5.5 2.6 4.4 1.2 1
6.1 3.0 4.6 1.4 1
5.8 2.6 4.0 1.2 1
5.0 2.3 3.3 1.0 1
5.6 2.7 4.2 1.3 1
5.7 3.0 4.2 1.2 1
5.7 2.9 4.2 1.3 1
6.2 2.9 4.3 1.3 1
5.1 2.5 3.0 1.1 1
5.7 2.8 4.1 1.3 1
6.3 3.3 6.0 2.5 2
5.8 2.7 5.1 1.9 2

CLUSTERING ANALYSIS

- Clustering analysis aims to find centers of clusters and memberships
- Criterion is the mean distance between data points and correspondent centers

$$E = \frac{1}{N} \sum_i \min_j \|x_i - y_j\|$$

CLUSTERING ANALYSIS

```
function R=demo_iris()
Z = load('iris.dat');
[N d]=size(Z);
X=Z(:,1:d-1);Y=Z(:,d)+1;
K=7;
[C Q]=annealed_kmeans2(X,K);
[ss ind]=max(Q');
R=zeros(K,max(Y));
for t=1:length(ind)
    i=ind(t);j=Y(t);
    R(i,j)=R(i,j)+1;
end
```

Z = load('iris.txt');

Y
collects
labels

[N d]=size(Z);
X=Z(:,1:d-1);Y=Z(:,d)+1;
K=7;

[C Q]=annealed_kmeans2(X,K);

X
collects
features

C
collects K
centers

Q collects
memberships

RESULTS I: CLUSTER-LABEL MATRIX

```
[ss ind]=max(Q');  
R=zeros(K,max(Y));  
E=0;
```

R(i,j) denotes
the pattern
number of
type j in the
ith cluster

```
for t=1:length(ind)
```

```
i=ind(t);j=Y(t);  
R(i,j)=R(i,j)+1;  
E=E+1/N*sqrt(sum(((X(t,:)-C(j,:)).^2));
```

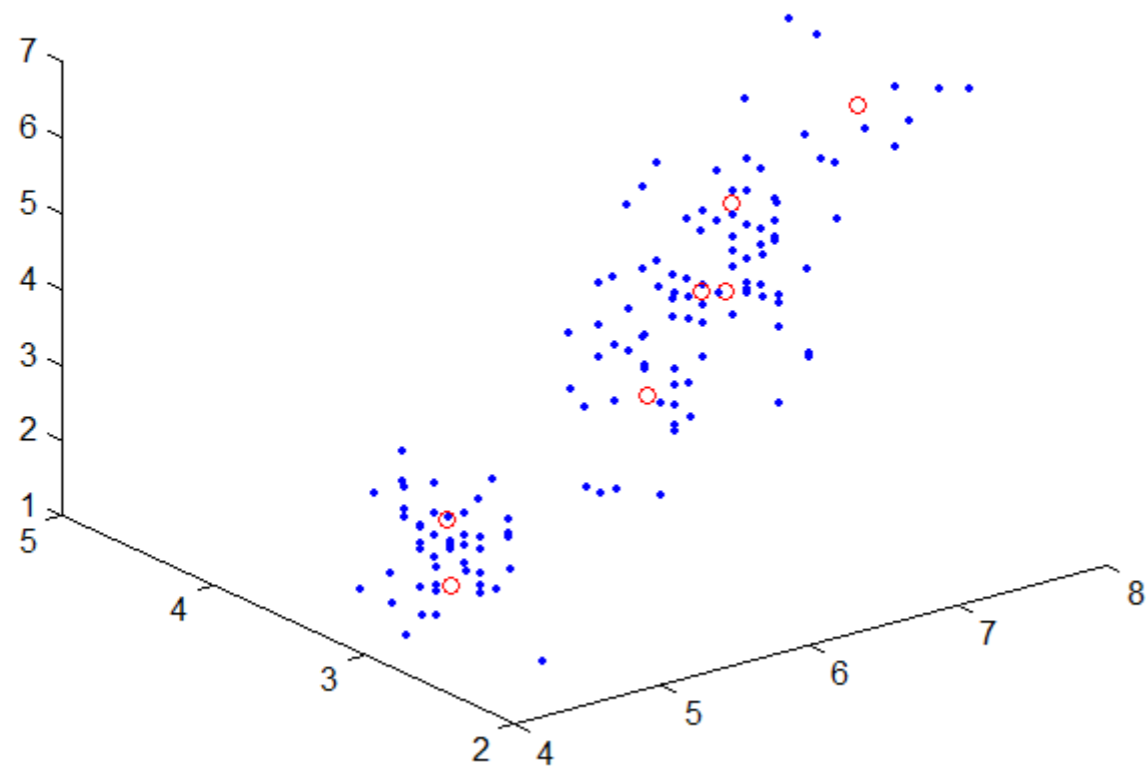
ans =

0	0	12
0	26	1
0	21	0
26	0	0
24	0	0
0	3	15
0	0	22

$$E = \frac{1}{N} \sum_i \min_j \|x_i - y_j\|$$

RESULT 2

feature 1 2 3



feature 1 2 4

