Fuzzy Cluster Membership

Incorporating Fuzzy Cluster Memberships within SAS Enterprise Miner

Donald K. Wedding, PhD
Principal Industry Consultant – Advanced Analytics
SAS Institute

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

- **What is Clustering**
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

What Is Clustering

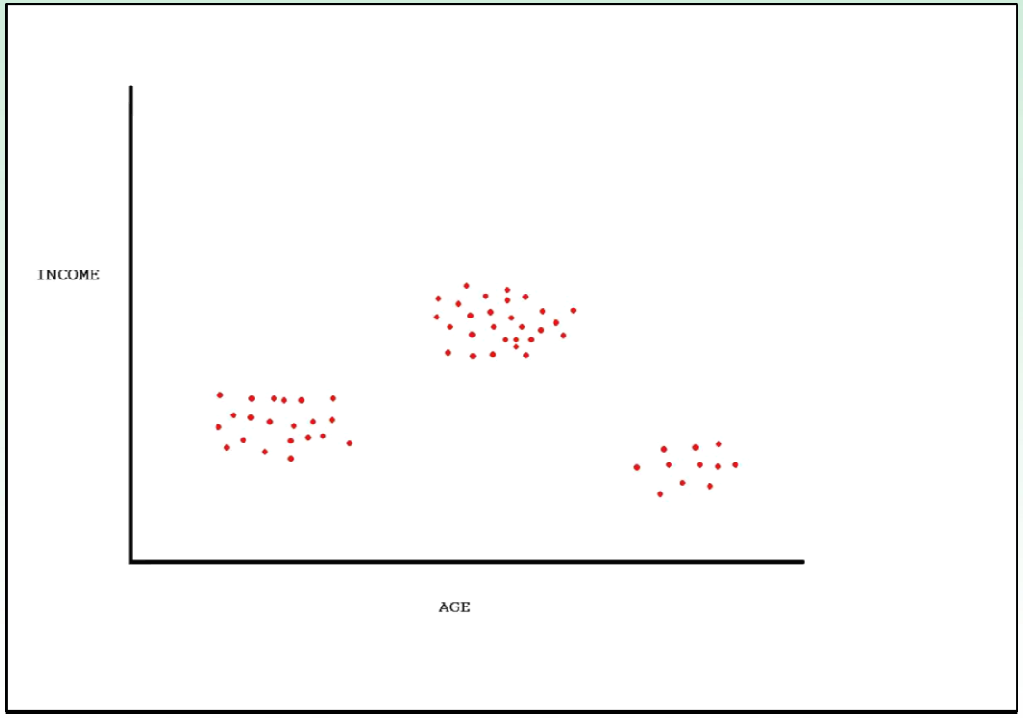
Process of placing data into groups so that...

- Members IN the group is SIMILAR
- Members NOT in the group are DIFFERENT

What Is Clustering

How many clusters do you see?

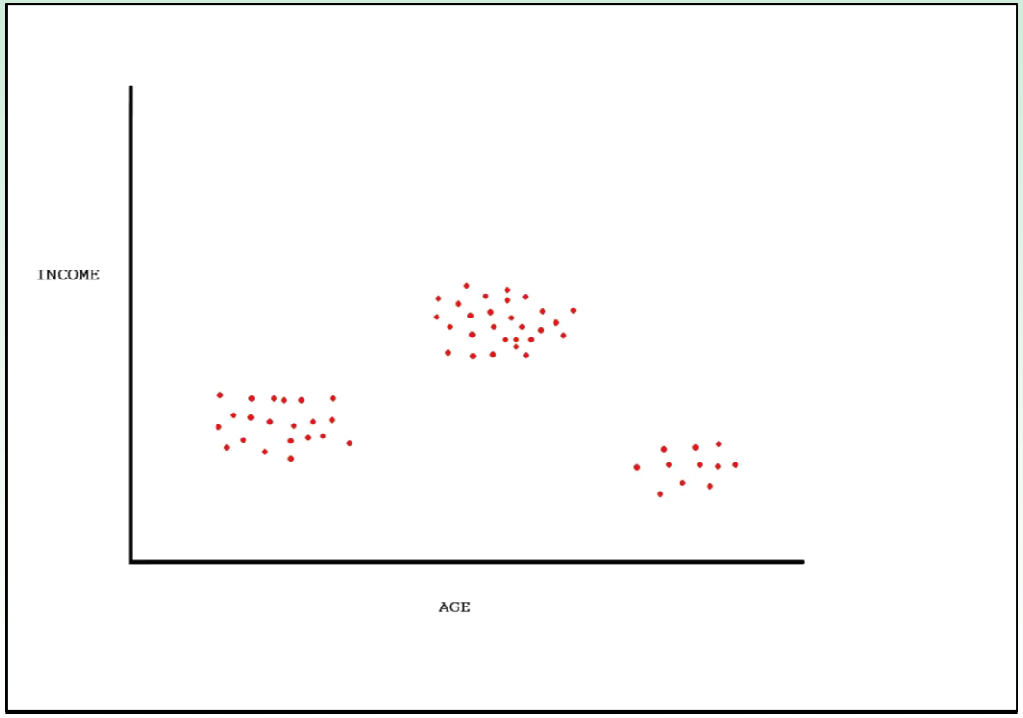
(Hint: This is not a trick question)



What Is Clustering

Visual Inspection "proc eyeball"

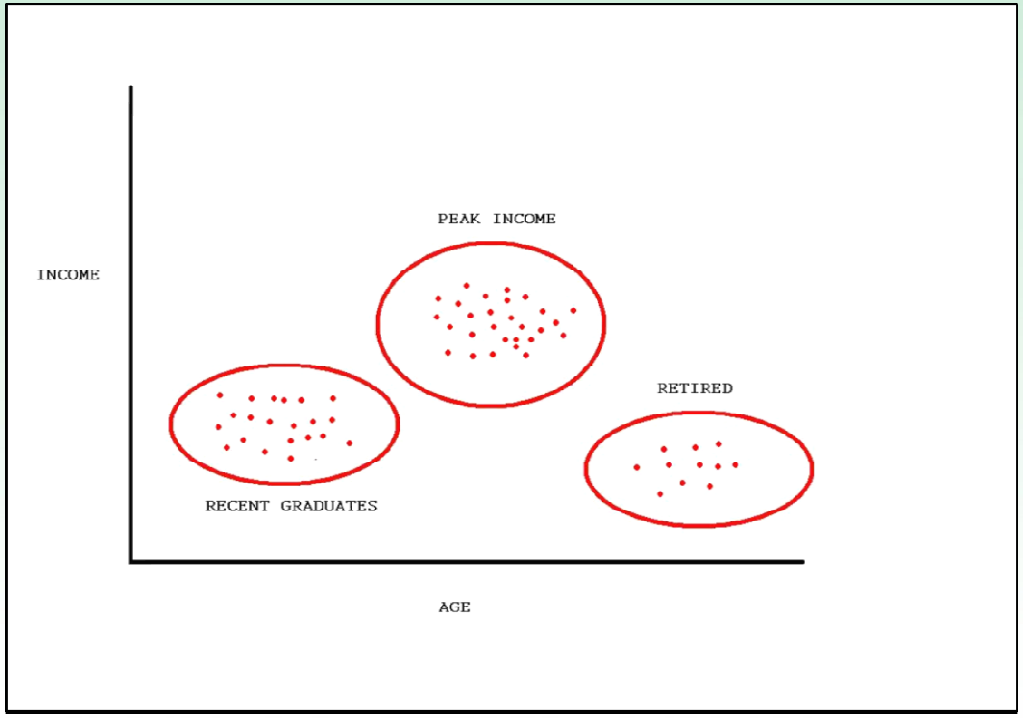
There are **three** clusters.



What Is Clustering

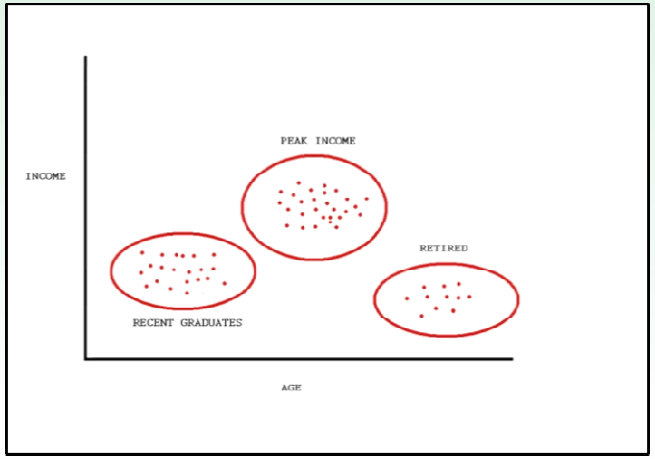
Visual Inspection "proc eyeball"

There are **three** clusters.



What Is Clustering

- A bank might use these clusters for “cross sell”
  - RECENT GRADUATES : Overdraft Protection
  - PEAK INCOME : Mortgage and Investment Accounts
  - RETIRED : IRA, Trust Accounts



- What is Clustering
- **How is Data Clustered**
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

How is Data Clustered

- Most data has more than two dimensions, for example: (Age, Income, Wealth, Deposit Balance)
  - Visual Inspection ("proc eyeball") is not possible
  - Algorithms are used.

- Most data mining software has two algorithms
  - K-Means Clustering
  - Kohonen / SOM Neural Networks

- What is Clustering
- How is Data Clustered
  - **K Means Clustering**
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

K-Means Clustering

- Iterative clustering algorithm.
- Widely used because:
  - Fast Algorithm
  - Guaranteed to converge
- Solution is not optimal, but is usually good.

K-Means Clustering

1. For N data points, randomly generate k cluster centers where k<N
2. Calculate, the distance (usually Euclidean) from each point to each cluster center.
3. Place each record in nearest cluster.
4. Find new cluster centers (usually the mean value of the members assigned to that cluster.
5. Go to Step 2, repeat until cluster membership no longer changes.

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - **Kohonen/SOM Clustering**
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Kohonen (SOM) Clustering

- (SOM=Self Organizing Map)

- Similar to K-Means but with differences:
  - Clusters are in a grid
  - Each cluster competes to acquire records.
    - Winning clusters are rewarded
    - Clusters next to the winner are rewarded
  - Adjacent clusters will be similar to one another.

Kohonen (SOM) Clustering

1. For N data points, select the desired dimensions of SOM (PxQ) where PxQ<N
2. Randomly select starting points of each cluster
3. *Competition:* Calculate, the distance (usually Euclidean) from each point to each cluster center. Place each record in nearest cluster.
4. *Adaption*: Reward the winning cluster by adjusting its center point to move closer to the record.
5. *Cooperation*: Move the cluster centers of the adjacent clusters so that they also move closer to the record.
6. Go to Step 3, repeat until cluster membership no longer changes.

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - **Advantages and Disadvantages of K-Means and Kohonen/SOM**
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

K-Means / Kohonen (SOM)

Advantages

- Algorithms are easy to implement for programmers
- Simple to use
- Quick to Converge
- Usually Give Good Results (Explain Data, Predictive Results)
- Widely Used

K-Means / Kohonen (SOM)

Disadvantage

- Winner Take All Algorithms
  - Also known as … HARD CLUSTERS

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- **Difficulties with Hard Clusters**

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

HARD CLUSTERING

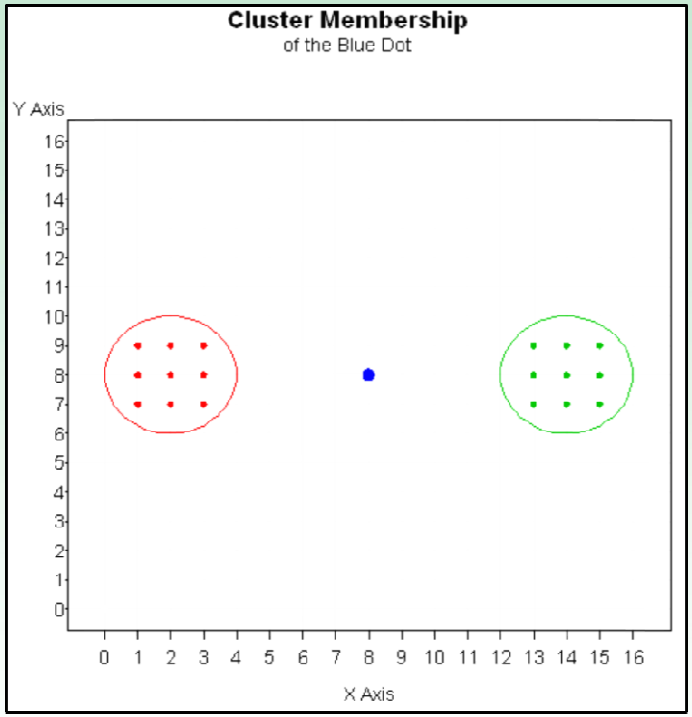
- A record can be in 1 and only 1 cluster
- “Winner Take All”

HARD CLUSTERING

- Problems with HARD CLUSTERING
  - What happens when a record is equally close to two clusters?
  - What happens when data is NOT compact and well separated?

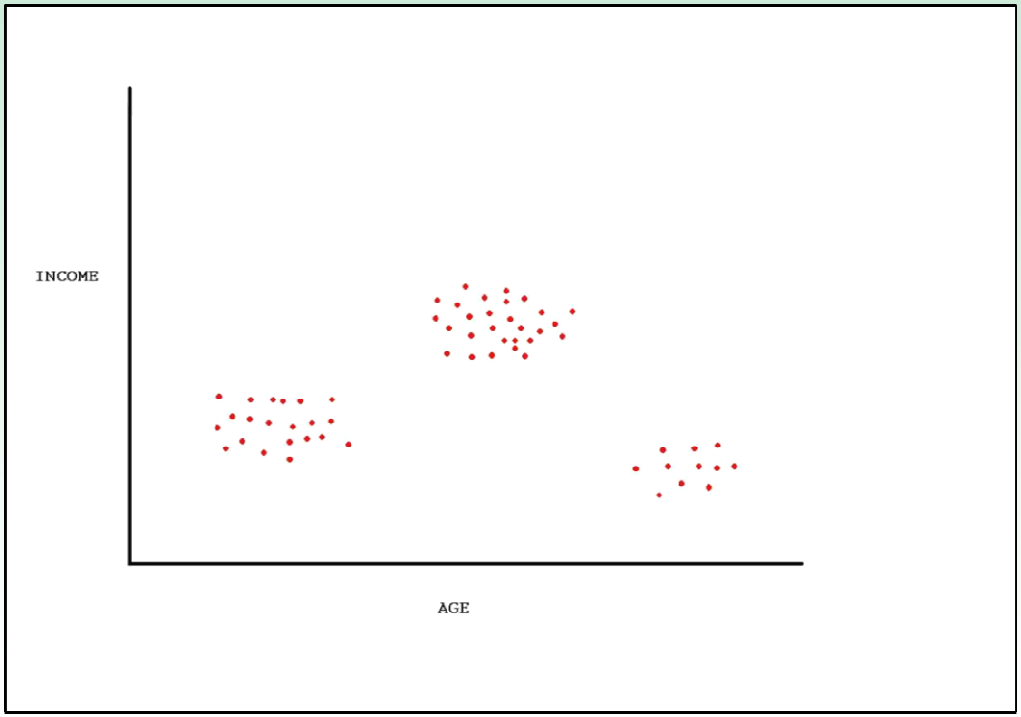
HARD CLUSTERING (equal distance)

Is the BLUE dot in the RED cluster or the GREEN cluster?



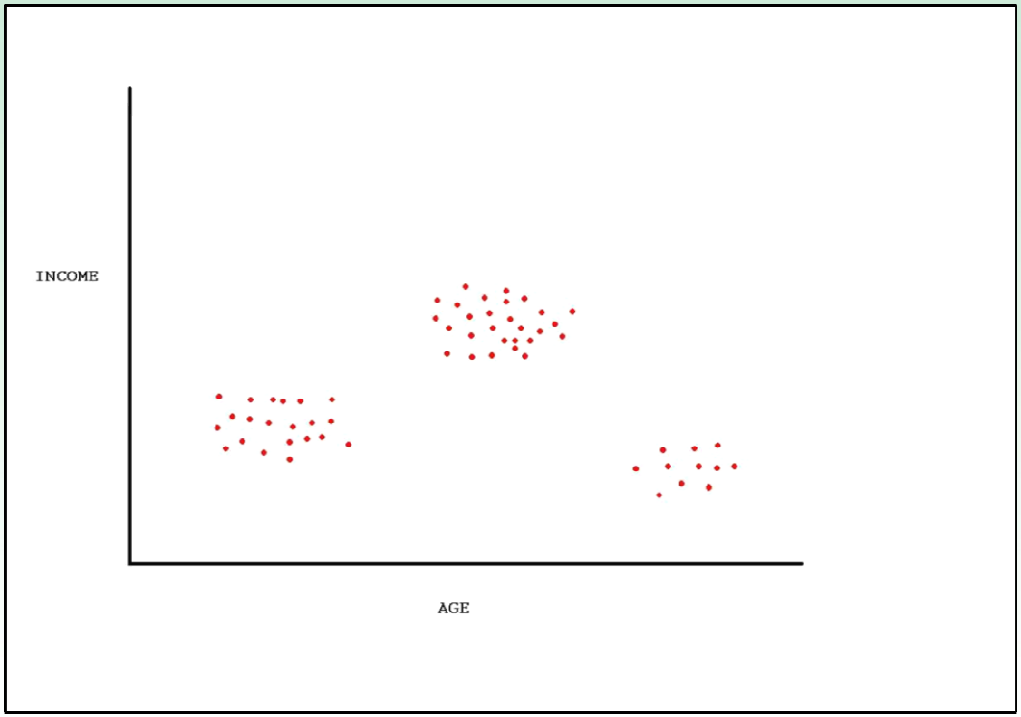
HARD CLUSTER (easy when…)

- Data COMPACT and WELL SEPARATED
  - How many clusters are there?



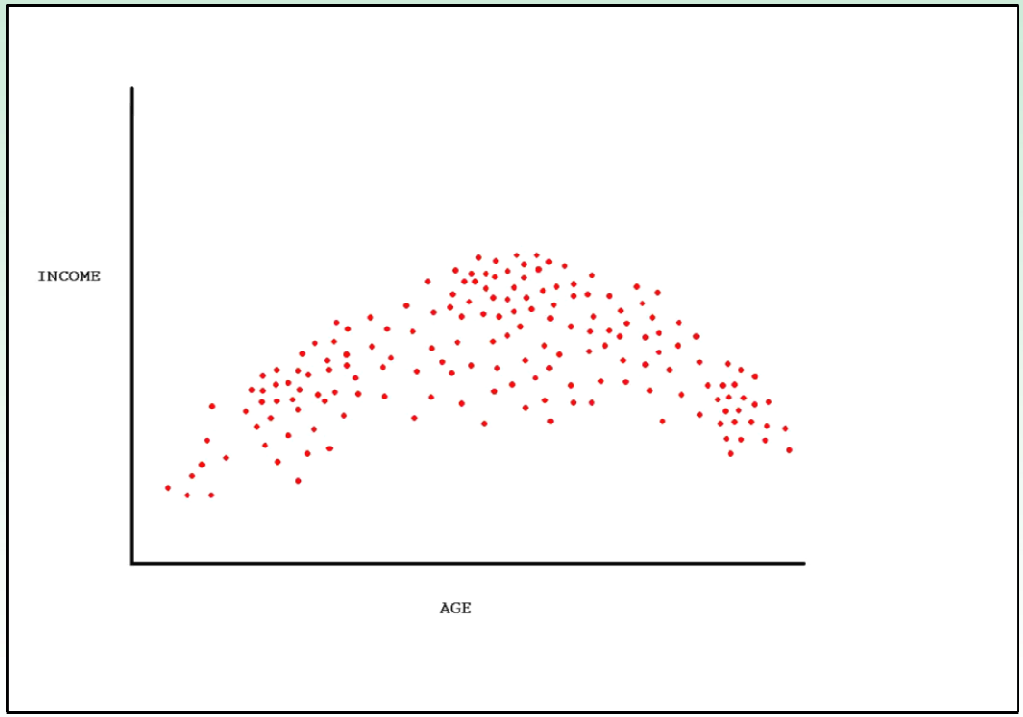
HARD CLUSTER (difficult when…)

- Data NOT COMPACT and NOT WELL SEPARATED



HARD CLUSTER (difficult when…)

- Data NOT COMPACT and NOT WELL SEPARATED
  - How many clusters are there?



- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- **Fuzzy Clustering**
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

FUZZY CLUSTERING

- Extension of Fuzzy Set Theory
- A record can be PARTIALLY in a Cluster
- A record can be in MORE THAN ONE Cluster
- Total Cluster Memberships Sum to 1
- *Hard Clusters* are a special case of *Fuzzy Clusters*

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - **Fuzzy Logic**
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy Logic

- Described by Lotfi Zadeh in 1965
- Extends Boolean Logic (True,False) (1,0)
- Allows a fact to be PARTIALLY TRUE
- Membership functions must be estimated.
  - (Usually by an expert in the field)

Fuzzy Logic (example)

- In the NBA, a player is usually considered “Tall” if they are 7 feet tall.
  - **Yao Ming** is 7’6”
    - TALL
  - **Spud Webb** 5’6”
    - SHORT … 3rd shortest player ever

- Question
  - **Tim Duncan** is 6’11”
    - TALL or SHORT?

Fuzzy Logic (example pg.2)

- In Boolean, **Tim Duncan** at 6'11" would be…

  SHORT

- In Fuzzy Logic, **Tim Duncan** at 6'11" would be…

  95% TALL

  5% SHORT

  (Assume membership provided by expert)

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - **Fuzzy C-Means Algorithm**
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy Clustering

What is Fuzzy Clustering?

Fuzzy Clustering

- Extension of Zadeh's Fuzzy Set Theory
- Allows a record to be in more than one cluster by assigning partial membership to each cluster.
- Many algorithms, but most widely used is Fuzzy C-Means described by Bezdek (1981).

Fuzzy Clustering

**50% Membership
in each cluster**

**Cool and
Not Smart**

**Cool and
Smart**

**Smart and
Not Cool**

Fuzzy Clustering

**50% Membership
in each cluster**

**Cool and
Not Smart**

**Not Cool and
Not Smart**

**Smart and
Not Cool**

K-Means Clustering

1. For N data points, randomly generate k cluster centers where k<N
2. Calculate, the distance (usually Euclidean) from each point to each cluster center.
3. Place each record in nearest cluster.
4. Find new cluster centers (usually the mean value of the members assigned to that cluster.
5. Go to Step 2, repeat until cluster membership no longer changes.

Fuzzy C-Means Clustering

1. For N data points, randomly generate k cluster centers where k<N

2. Calculate, the distance (usually Euclidean) from each point to each cluster center.

3. Assign a ***proportional*** or ***fuzzy membership*** for each of the N records to each of the k clusters.

4. Find new cluster centers (usually the weighted mean value of the members assigned to that cluster.

5. Go to Step 2, repeat until cluster membership no longer changes (or until some other convergence criteria is met).

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - **Fuzzy Membership Function**
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy Membership Function

- The membership function is given as:

$$u_k = \frac{1}{\sum_{i=1}^{j} \left(\frac{d_k}{d_i} \right)^p}$$

- u_k is the membership of the record in cluster k
- j is the total number of clusters
- d_k is the distance from the record to cluster k
- d_i is the distance from the record to cluster i
- p is the fuzzy exponent

Fuzzy Membership Function Distance Metric

- Distance from a record to a center
  - Any metric is valid, so long as it follows the rules proposed by Frechet (1906).
    - D(X,X) = 0 (Identity)
    - D(X,Y) >= 0 (Non negative)
    - D(X,Y) = D(Y,X) (No one way streets)
    - D(X,Y) <= D(X,Z) + D(Z,Y) (Shortest Distance is a line)

Fuzzy Membership Function Distance Metric

- Most popular distance metric is Euclidean

$$d = \sqrt{\left(\sum_{i=i}^{N} |x_i - y_i|^2\right)}$$

Fuzzy Membership Function
Fuzzy Exponent

- The value *p* (fuzzy exponent) is calculated as:

$$p = \frac{2}{(m-1)}$$

- *m* is a constant set by the user where 1<m<∞

•	*as*	*m → 1*	then	p → ∞ (Results approach "hard clusters")
•	*for*	*m = 2*	then	p=2 (Usually this value is used)
•	*for*	*m = 3*	then	p=1 (Linear relationship)
•	*as*	*m → ∞*	then	p → 0 ("All memberships the same")
•	Typically			1<m<3
•	Usually			m=2

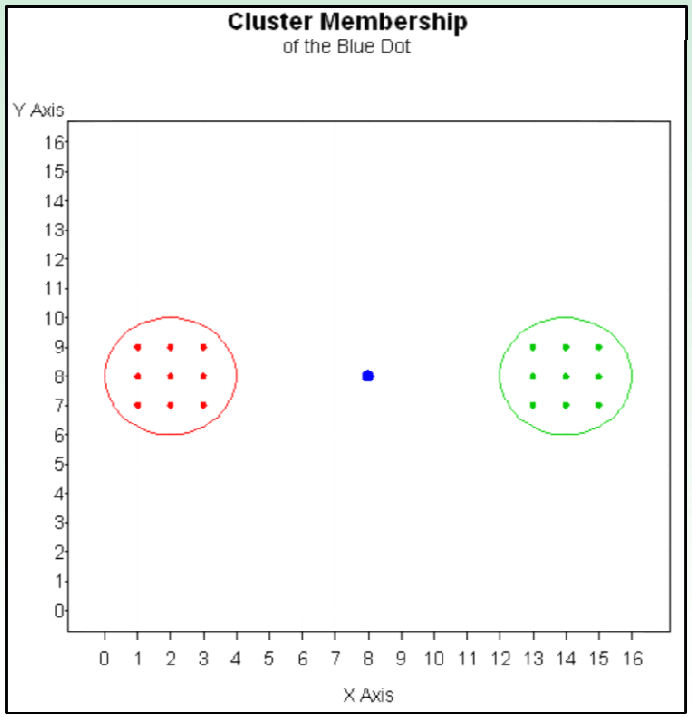
- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - **Example Of Fuzzy Membership**
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

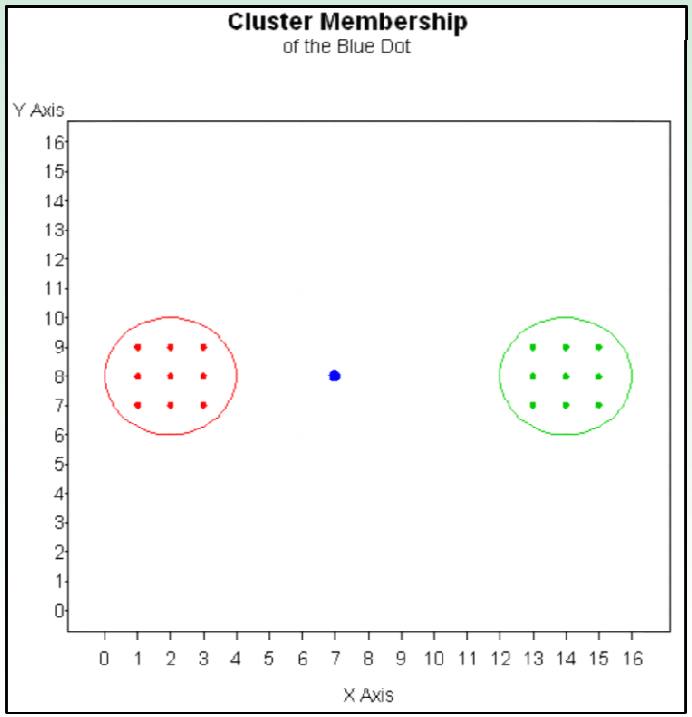
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	6	?	0.500	0.500	0.500	0.500
GREEN	6	?	0.500	0.500	0.500	0.500



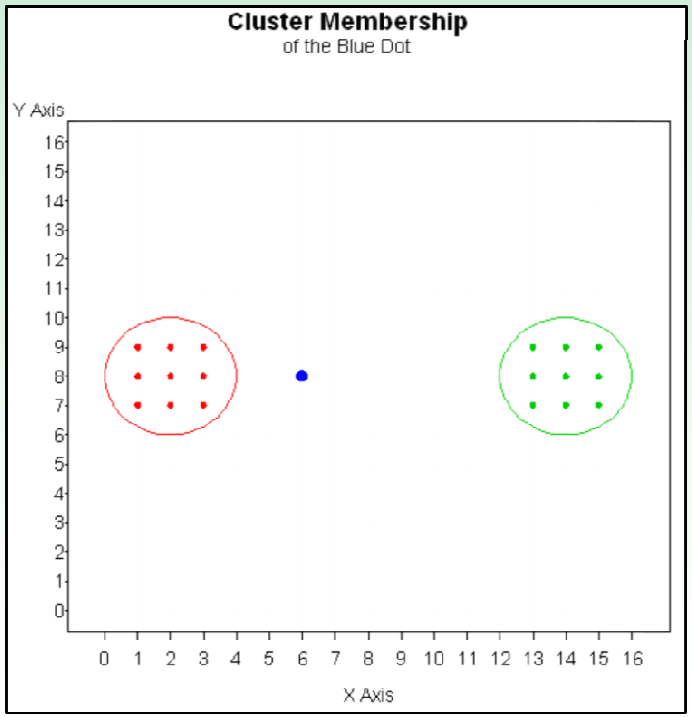
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	5	1	0.793	0.662	0.583	0.502
GREEN	7	0	0.207	0.338	0.417	0.498



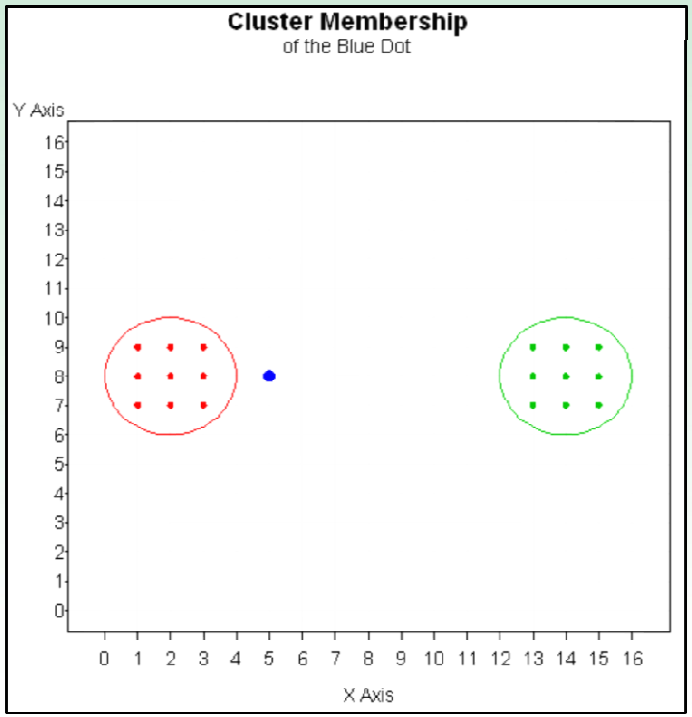
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	4	1	0.941	0.800	0.667	0.504
GREEN	8	0	0.059	0.200	0.333	0.496



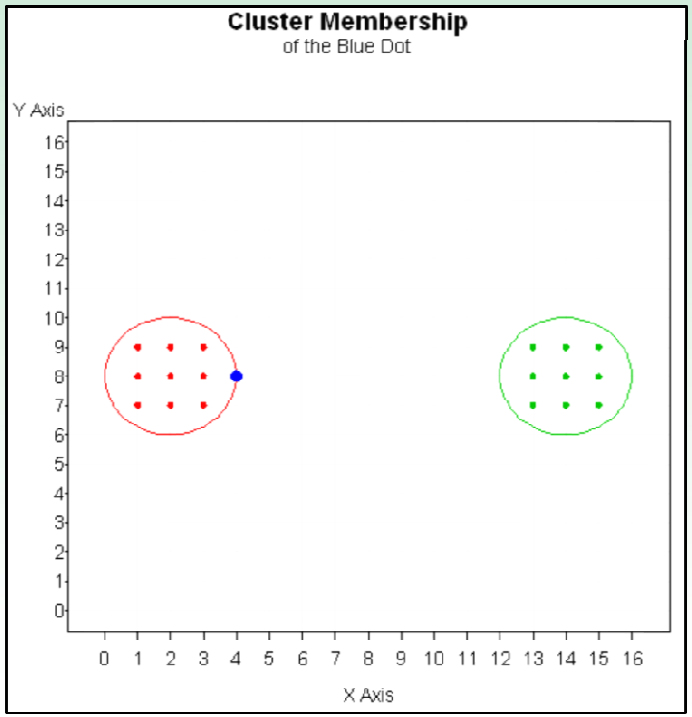
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	3	1	0.988	0.900	0.750	0.506
GREEN	9	0	0.012	0.100	0.250	0.494



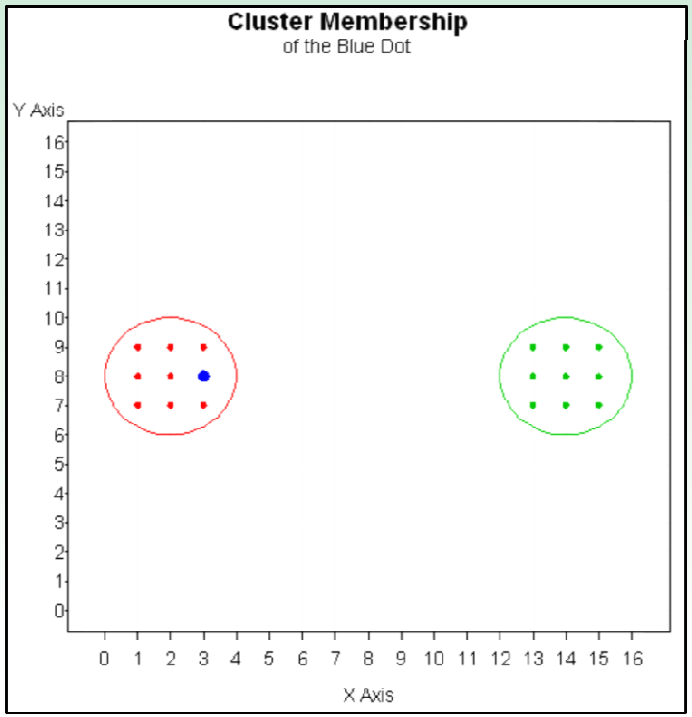
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	2	1	0.998	0.962	0.833	0.508
GREEN	10	0	0.002	0.038	0.167	0.492



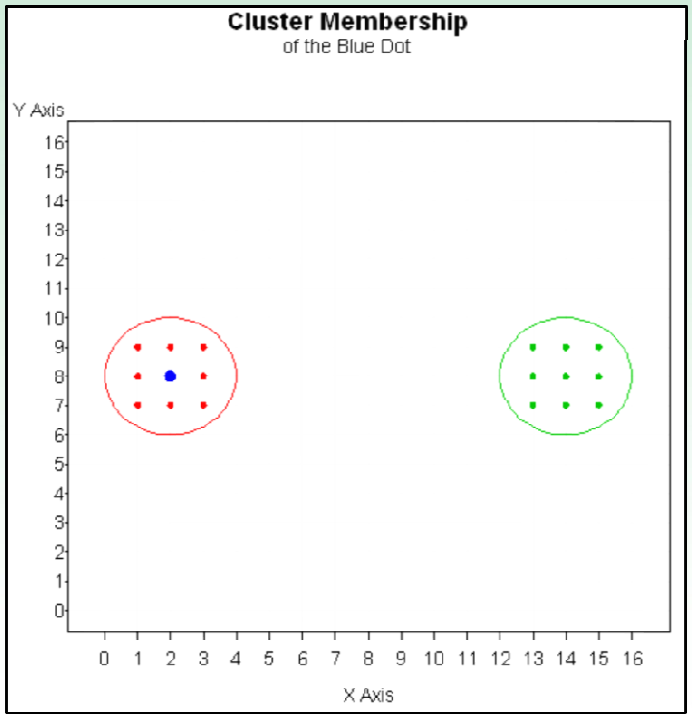
Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	1	1	1	0.992	0.917	0.512
GREEN	11	0	0	0.008	0.083	0.488



Example
Membership in Red Cluster

CLUSTER	DISTANCE	HARD CLUSTER	M = 1.5	M=2.0	M=3.0	M=100
RED	0	1	1	1	1	1
GREEN	12	0	0	0	0	0



Fuzzy C-Means Clustering

U_{red} for different *m* values

D_{red}	D_{green}	Hard	1.1	1.5	2	3	1000
6	6	?	0.500	0.500	0.500	0.500	0.500
5	7	1	0.999	0.793	0.662	0.583	0.500
4	8	1	1.000	0.941	0.800	0.667	0.500
3	9	1	1.000	0.988	0.900	0.750	0.501
2	10	1	1.000	0.998	0.962	0.833	0.501
1	11	1	1.000	1.000	0.992	0.917	0.501
0	12	1	1.000	1.000	1.000	1.000	1.000
100	101	1	0.550	0.510	0.505	0.502	0.500
1000	1001	1	0.505	0.501	0.500	0.500	0.500

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - **Advantages and Disadvantages of Fuzzy Clustering**

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy C-Means Clustering

Advantages

- More generalized form of Hard Clusters.
  - Handles data that equidistant to multiple clusters
  - Handles data that is not compact and well separated
- Algorithms are easy to implement for programmers
- Simple to use (but not as simple as hard clustering)
- Usually Give Good Results

Fuzzy C-Means Clustering

Disadvantage

- Can be slow to converge
- What value should be used for *m* ?
- Commercially available data mining software does not usually implement fuzzy clustering.

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- **Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)**
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy Cluster Approximation

Question

- If only **hard clusters** are available, how can **fuzzy clusters** be used?

Fuzzy Cluster Approximation

Requirements

1. Cluster Centers

2. Membership Function

Fuzzy Cluster Approximation

Cluster Centers

1. Commercial software implements HARD clustering techniques.
   - K-Means
   - Kohonen Networks / SOM
2. Use these techniques to find HARD cluster centers.

Fuzzy Cluster Approximation

Fuzzy Membership

Calculate the fuzzy memberships using the hard cluster centers and the formula:

$$u_k = \frac{1}{\sum_{i=1}^{j} \left(\frac{d_k}{d_i} \right)^p}$$

Fuzzy Cluster Approximation

Fuzzy Membership

1. Not exactly the same as Bezdek's Fuzzy C Means algorithm, only an approximation.

2. Doesn't matter. It's useful.

3. *"Every model is wrong, but some models are useful" – George E.P. Box*

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- **Fuzzy Cluster Approximation in Enterprise Miner**
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Fuzzy Cluster Approximation
SAS Enterprise Miner

1. Find Cluster Centers With:
   - K-Means
   - Kohonen/Som

2. Copy the SAS Score code to an EM SAS Code Node

3. Add a call to a macro program to calculate Fuzzy Membership

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - **Fuzzy Membership Macro**
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

FUZZY CLUSTER MACRO (page 1)

```
%macro fuzzyCluster( DARRAY, UARRAY, SIZE, M=2 );

array &UARRAY.  &UARRAY.1-&UARRAY.&SIZE.;
array D&UARRAY.[&SIZE.] _temporary_;

length      POWER 8.;               drop        POWER;
length      DISTSUM 8.;             drop        DISTSUM;
length      DISTFLAG 8.;            drop        DISTFLAG;
length      i 8.;                   drop        i;
length      MAXD 8.;                drop        MAXD;

POWER = 2/(&M.-1);

MAXD = -1;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = sqrt(&DARRAY.[i]);
     if D&UARRAY.[i] > MAXD then MAXD = D&UARRAY.[i];
end;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = D&UARRAY.[i] / MAXD;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
     DISTSUM = DISTSUM + D&UARRAY.[i]**POWER;
end;
```

FUZZY CLUSTER MACRO (page 2)

```
do i = 1 to &SIZE.;
   &UARRAY.[i] = 1 / ( (D&UARRAY.[i]**POWER)*DISTSUM  );
end;

DISTFLAG = 0;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then DISTFLAG = 1;
end;

if DISTFLAG > 0 then do;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then
      &UARRAY.[i] = 1;
   else
      &UARRAY.[i] = 0;
end;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
   DISTSUM = DISTSUM + &UARRAY.[i];
end;

do i = 1 to &SIZE.;
   &UARRAY.[i] = &UARRAY.[i] / DISTSUM;
   &UARRAY.[i] = round( &UARRAY.[i], 0.001 );
end;

%mend;
```

FUZZY CLUSTER MACRO (page 1)

```
%macro fuzzyCluster( DARRAY, UARRAY, SIZE, M=2 );

array &UARRAY.  &UARRAY.1-&UARRAY.&SIZE.;
array D&UARRAY.[&SIZE.] _temporary_;

length      POWER 8.;                drop        POWER;
length      DISTSUM 8.;              drop        DISTSUM;
length      DISTFLAG 8.;             drop        DISTFLAG;
length      i 8.;                    drop        i;
length      MAXD 8.;                 drop        MAXD;

POWER = 2/(&M.-1);

MAXD = -1;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = sqrt(&DARRAY.[i]);
     if D&UARRAY.[i] > MAXD then MAXD = D&UARRAY.[i];
end;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = D&UARRAY.[i] / MAXD;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
     DISTSUM = DISTSUM + D&UARRAY.[i]**POWER;
end;
```

Distance Array
(already calculated by Cluster Node)

Fuzzy Array
(Calculated by macro)

Size of the arrays
(How many clusters are there?)

Fuzzy Factor
(Set by user, defaults to 2)

FUZZY CLUSTER MACRO (page 1)

```
%macro fuzzyCluster( DARRAY, UARRAY, SIZE, M=2 );

array &UARRAY.  &UARRAY.1-&UARRAY.&SIZE.;
array D&UARRAY.[&SIZE.] _temporary_;

length      POWER 8.;               drop        POWER;
length      DISTSUM 8.;             drop        DISTSUM;
length      DISTFLAG 8.;            drop        DISTFLAG;
length      i 8.;                   drop        i;
length      MAXD 8.;                drop        MAXD;

POWER = 2/(&M.-1);

MAXD = -1;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = sqrt(&DARRAY.[i]);
     if D&UARRAY.[i] > MAXD then MAXD = D&UARRAY.[i];
end;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = D&UARRAY.[i] / MAXD;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
     DISTSUM = DISTSUM + D&UARRAY.[i]**POWER;
end;
```

Normalize all distances to 1

FUZZY CLUSTER MACRO (page 1)

```
%macro fuzzyCluster( DARRAY, UARRAY, SIZE, M=2 );

array &UARRAY.  &UARRAY.1-&UARRAY.&SIZE.;
array D&UARRAY.[&SIZE.] _temporary_;

length      POWER 8.;               drop        POWER;
length      DISTSUM 8.;             drop        DISTSUM;
length      DISTFLAG 8.;            drop        DISTFLAG;
length      i 8.;                   drop        i;
length      MAXD 8.;                drop        MAXD;

POWER = 2/(&M.-1);

MAXD = -1;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = sqrt(&DARRAY.[i]);
     if D&UARRAY.[i] > MAXD then MAXD = D&UARRAY.[i];
end;

do i = 1 to &SIZE.;
     D&UARRAY.[i] = D&UARRAY.[i] / MAXD;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
     DISTSUM = DISTSUM + D&UARRAY.[i]**POWER;
end;
```

Start of Bezdek's Fuzzy Membership Function

FUZZY CLUSTER MACRO (page 2)

```
do i = 1 to &SIZE.;
   &UARRAY.[i] = 1 / ( (D&UARRAY.[i]**POWER)*DISTSUM  );
end;

DISTFLAG = 0;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then DISTFLAG = 1;
end;

if DISTFLAG > 0 then do;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then
      &UARRAY.[i] = 1;
   else
      &UARRAY.[i] = 0;
end;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
   DISTSUM = DISTSUM + &UARRAY.[i];
end;

do i = 1 to &SIZE.;
   &UARRAY.[i] = &UARRAY.[i] / DISTSUM;
   &UARRAY.[i] = round( &UARRAY.[i], 0.001 );
end;

%mend;
```

Finish of Bezdek's Fuzzy Membership Function

FUZZY CLUSTER MACRO (page 2)

```
do i = 1 to &SIZE.;
   &UARRAY.[i] = 1 / ( (D&UARRAY.[i]**POWER)*DISTSUM  );
end;

DISTFLAG = 0;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then DISTFLAG = 1;
end;

if DISTFLAG > 0 then do;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then
      &UARRAY.[i] = 1;
   else
      &UARRAY.[i] = 0;
end;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
   DISTSUM = DISTSUM + &UARRAY.[i];
end;

do i = 1 to &SIZE.;
   &UARRAY.[i] = &UARRAY.[i] / DISTSUM;
   &UARRAY.[i] = round( &UARRAY.[i], 0.001 );
end;

%mend;
```

Handle the special case of a record on a cluster center

FUZZY CLUSTER MACRO (page 2)

```
do i = 1 to &SIZE.;
   &UARRAY.[i] = 1 / ( (D&UARRAY.[i]**POWER)*DISTSUM  );
end;

DISTFLAG = 0;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then DISTFLAG = 1;
end;

if DISTFLAG > 0 then do;
do i = 1 to &SIZE.;
   if missing( &UARRAY.[i] ) then
      &UARRAY.[i] = 1;
   else
      &UARRAY.[i] = 0;
end;
end;

DISTSUM = 0;
do i = 1 to &SIZE.;
   DISTSUM = DISTSUM + &UARRAY.[i];
end;

do i = 1 to &SIZE.;
   &UARRAY.[i] = &UARRAY.[i] / DISTSUM;
   &UARRAY.[i] = round( &UARRAY.[i], 0.001 );
end;

%mend;
```

Make certain that memberships add up to 1.0

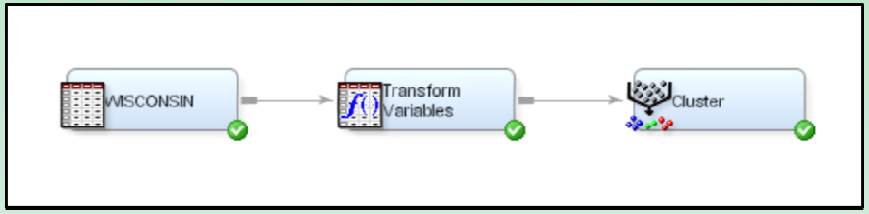
- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

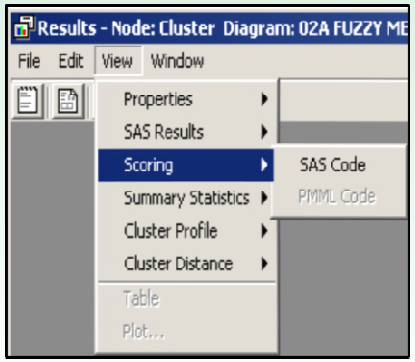
- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - **K-Means**
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

Approximating Fuzzy with K-MEANS

- STEP 1: Use a K-MEANS node to find Cluster Centers

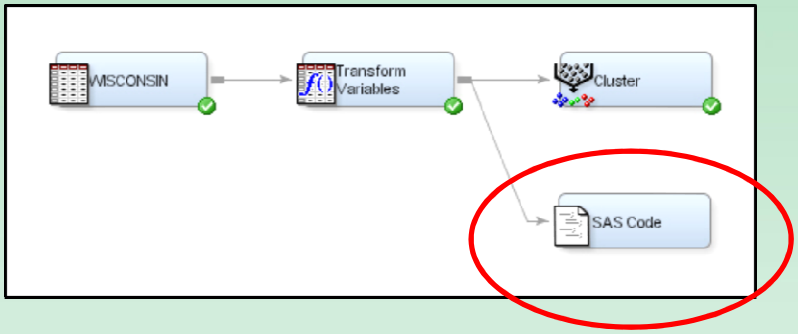


- STEP 2: Open the results window for the K-MEANS node and copy the SAS Score code to the Windows Buffer.



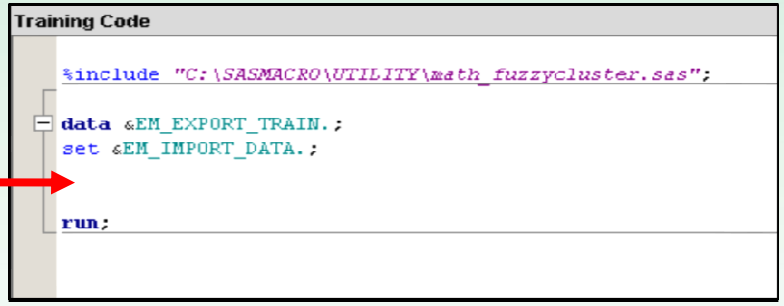
Approximating Fuzzy with K-MEANS

- STEP 3: Add a SAS Code node to the flow stream



- STEP 4: Write a blank data step inside of the SAS Code Node.

HERE !



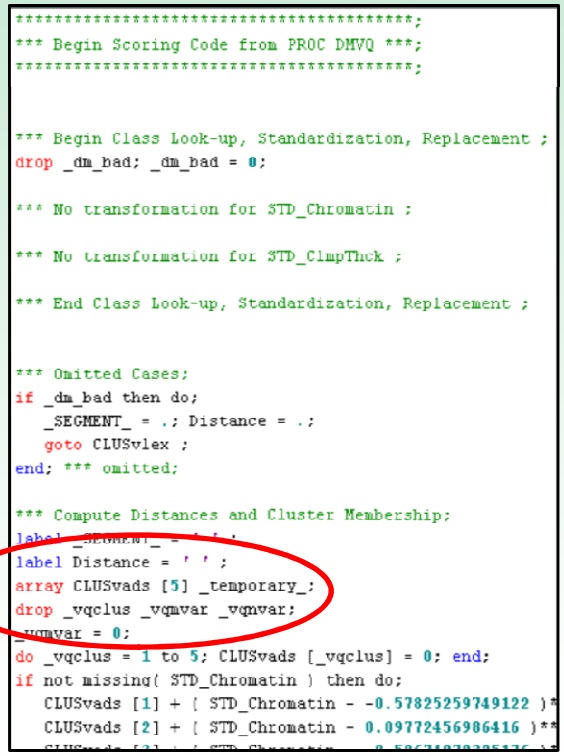
- STEP 5: Paste the SAS Score Code from **STEP 2** inside of the data step.

Approximating Fuzzy with K-MEANS

- STEP 6: Scroll to the top of the code in **STEP 5**. Locate the “distance” array near the top of program.

- STEP 7: Record the name of the array (in this case CLUSvads).

“CLUSvads” →



```
*****************************************;
*** Begin Scoring Code from PROC DMVQ ***;
*****************************************;


*** Begin Class Look-up, Standardization, Replacement ;
drop _dm_bad; _dm_bad = 0;

*** No transformation for STD_Chromatin ;

*** No transformation for STD_ClmpThck ;

*** End Class Look-up, Standardization, Replacement ;


*** Omitted Cases;
if _dm_bad then do;
   _SEGMENT_ = .; Distance = .;
   goto CLUSvlex ;
end; *** omitted;

*** Compute Distances and Cluster Membership;
label _SEGMENT_ = ' ' ;
label Distance = ' ' ;
array CLUSvads [5] _temporary_;
drop _vqclus _vqmvar _vqmvar;
_vqmvar = 0;
do _vqclus = 1 to 5; CLUSvads [_vqclus] = 0; end;
if not missing( STD_Chromatin ) then do;
   CLUSvads [1] + ( STD_Chromatin - -0.57825259749122 )*
   CLUSvads [2] + ( STD_Chromatin - 0.09772456906416 )**
```

Approximating Fuzzy with K-MEANS

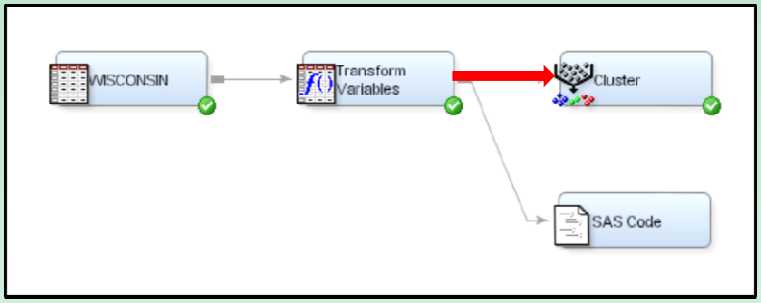
- STEP 8: Go to the end of the data step.
- STEP 9: Make the call to the SAS Macro program to calculate fuzzy membership.

• %fuzzyCluster	Name of the macro
• CLUSvads	Name of the array variable that holds the distances
• Ufuzzy	Name (given by user) to hold the fuzzy memberships
• SIZE = 5	User tells macro that there are 5 clusters
• M = 2	User tells macro to use an M value of 2

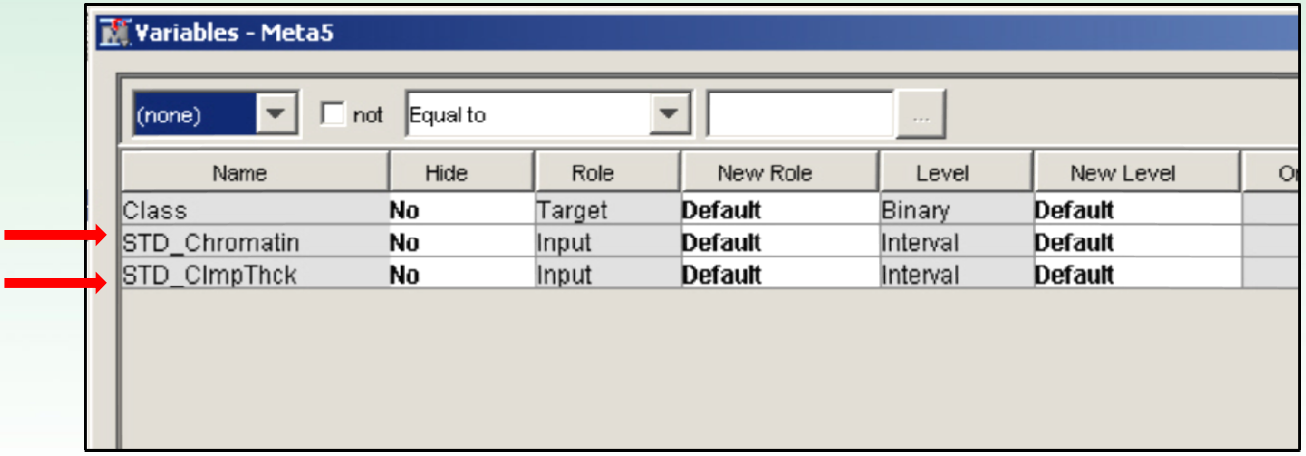
```
if _SEGMENT_ = 4 then _SEGMENT_LABEL_="Cluster4";
else
if _SEGMENT_ = 5 then _SEGMENT_LABEL_="Cluster5";

%fuzzyCluster( CLUSvads , Ufuzzy , SIZE=5, M=2 );

run;
```

Approximating Fuzzy with K-MEANS

- Variables Going INTO the K-MEANS Cluster Node

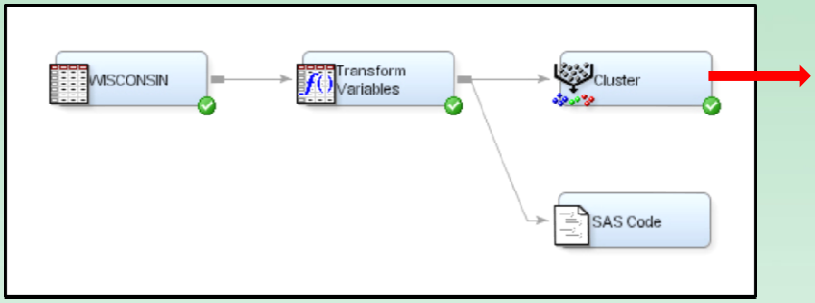


- K-Means is clustering on Standardized Chromatin and Clump Thickness



Approximating Fuzzy with K-MEANS

- Variables Going OUT of the K-MEANS Cluster Node



- K-Means creates a CLUSTER ("_SEGMENT_") and housekeeping variables.

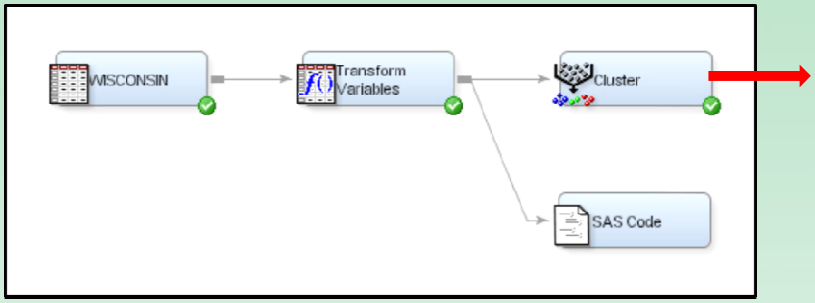
Variables - Meta3

(none) | not | Equal to

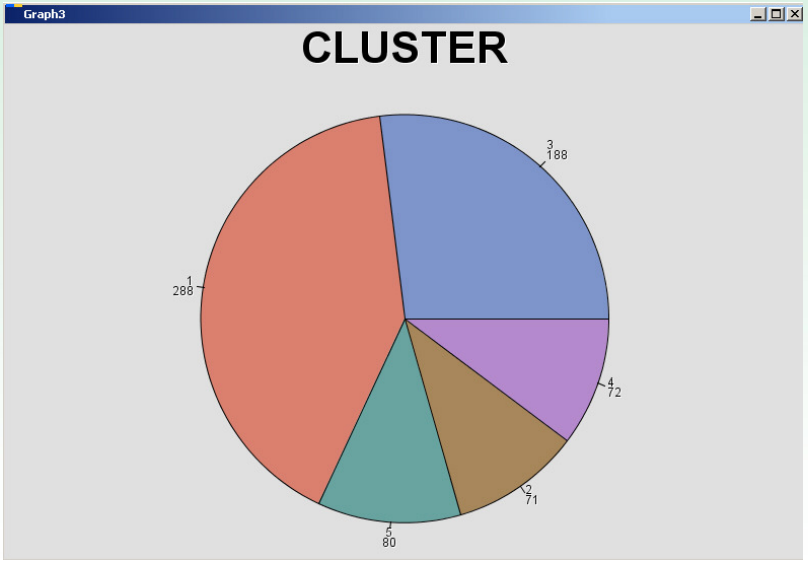
Name	Hide	Role	New Role	Level	New Level
Class	No	Target	Default	Binary	Default
Distance	No	Rejected	Default	Interval	Default
STD_Chromatin	No	Input	Default	Interval	Default
STD_ClmpThck	No	Input	Default	Interval	Default
SEGMENT	No	Segment	Default	Nominal	Default
_SEGMENT_LABEL_	No	Rejected	Default	Nominal	Default

Approximating Fuzzy with K-MEANS

- Variables Going OUT of the K-MEANS Cluster Node

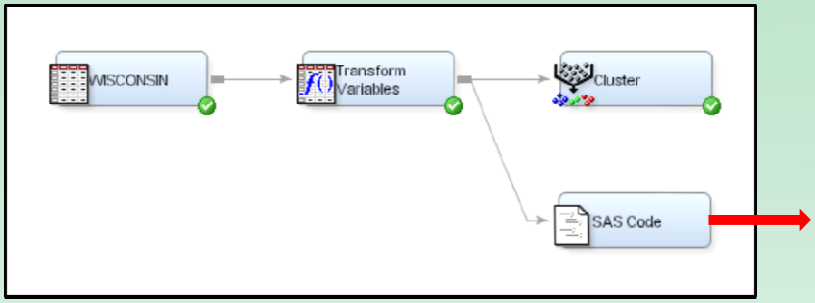


- Distribution of the clusters created by the K-Means Node



Approximating Fuzzy with K-MEANS

- Variables Going OUT of the FUZZY CLUSTER Code Node



- Same variables as K-Means Node … and also Fuzzy Memberships !

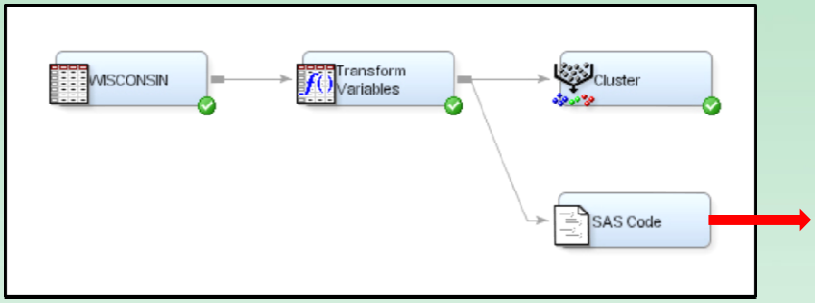
Variables - Meta6

(none) not Equal to

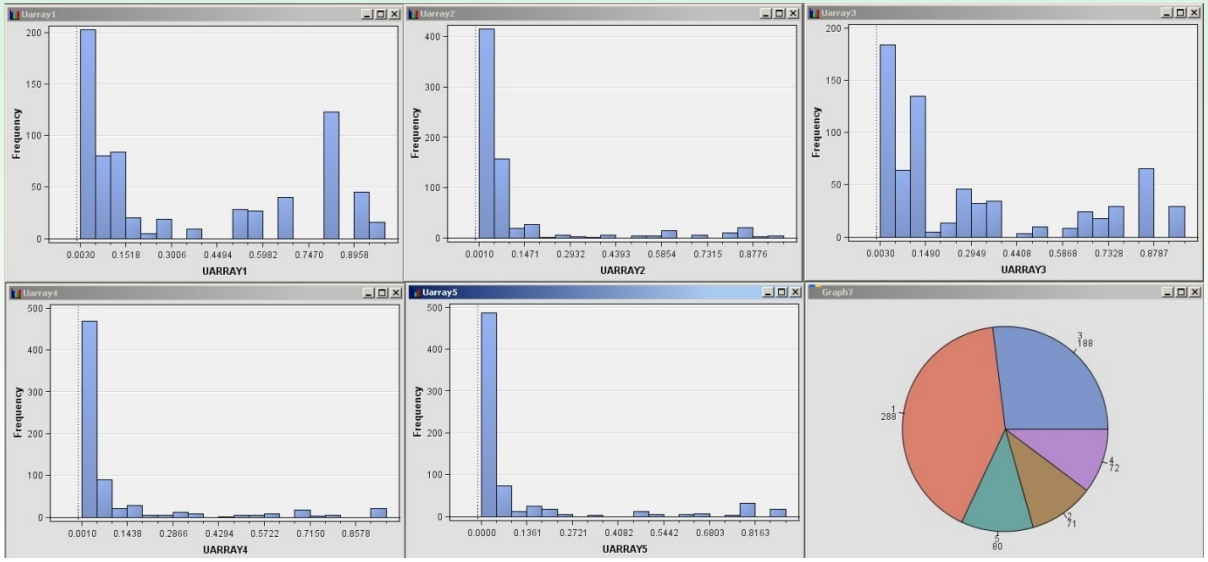
Name	Hide	Role	New Role	Level	New Level	Order	New Order
Class	No	Target	Default	Binary	Default		Default
Distance	No	Input	Default	Interval	Default		Default
STD_Chromat	No	Input	Default	Interval	Default		Default
STD_ClmpTh	No	Input	Default	Interval	Default		Default
Uarray1	No	Input	Default	Interval	Default		Default
Uarray2	No	Input	Default	Interval	Default		Default
Uarray3	No	Input	Default	Interval	Default		Default
Uarray4	No	Input	Default	Interval	Default		Default
Uarray5	No	Input	Default	Interval	Default		Default
SEGMENT	No	Segment	Default	Nominal	Default		Default
_SEGMENT_L	No	Input	Default	Nominal	Default		Default

Approximating Fuzzy with K-MEANS

- Variables Going OUT of the FUZZY CLUSTER Code Node



- Distribution of the clusters and fuzzy memberships created



APPROXIMATE FUZZY WITH K-MEANS

Add a call to the macro program

- Multiple calls to the macro are possible, so that the user may calculate different fuzzy memberships for different values of *m*.
- The variable &DIST is a macro variable holding the distance array. It is used for convenience.

```
if _SEGMENT_ = 4 then _SEGMENT_LABEL_="Cluster4";
else
if _SEGMENT_ = 5 then _SEGMENT_LABEL_="Cluster5";

%fuzzyCluster( &DIST. , Ua , SIZE=&ARRAYSIZE., M=1.01 );
%fuzzyCluster( &DIST. , Ub , SIZE=&ARRAYSIZE., M=1.10 );
%fuzzyCluster( &DIST. , Uc , SIZE=&ARRAYSIZE., M=1.25 );
%fuzzyCluster( &DIST. , Ud , SIZE=&ARRAYSIZE., M=1.30 );
%fuzzyCluster( &DIST. , Ue , SIZE=&ARRAYSIZE., M=1.35 );
%fuzzyCluster( &DIST. , Uf , SIZE=&ARRAYSIZE., M=1.40 );
%fuzzyCluster( &DIST. , Ug , SIZE=&ARRAYSIZE., M=2.00 );

run;
```

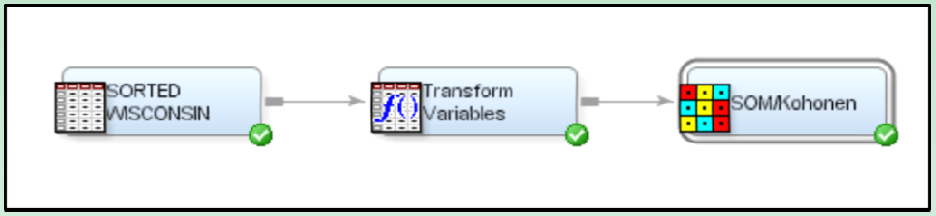
- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

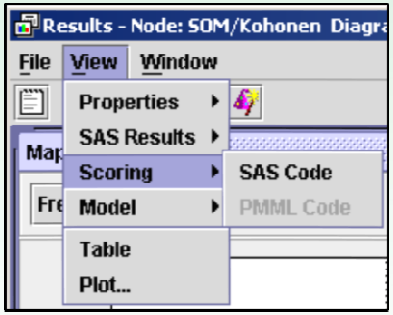
- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - **Kohonen / SOM Clustering**
- Example showing Fuzzy can improve accuracy
- Conclusion

Approximating Fuzzy with Kohonen/SOM

- STEP 1: Use a Kohonen/SOM node to find Cluster Centers

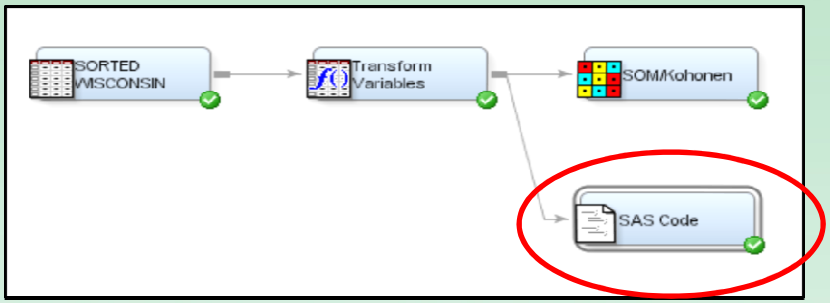


- STEP 2: Open the results window for the Kohonen/SOM node and copy the SAS Score code to the Windows Buffer.

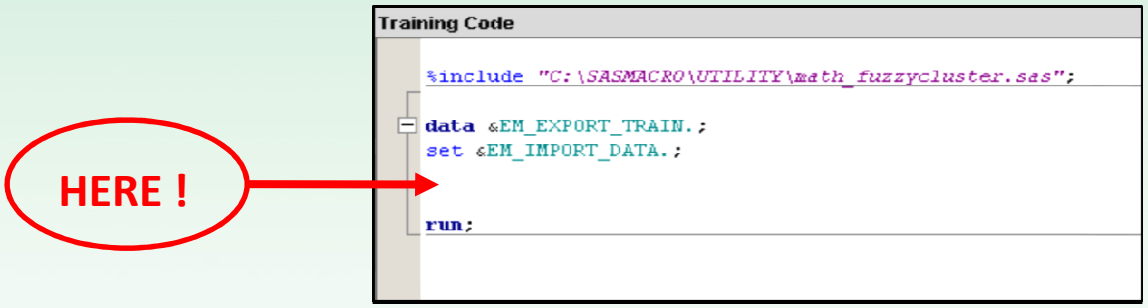


Approximating Fuzzy with Kohonen/SOM

- STEP 3: Add a SAS Code node to the flow stream



- STEP 4: Write a blank data step inside of the SAS Code Node.



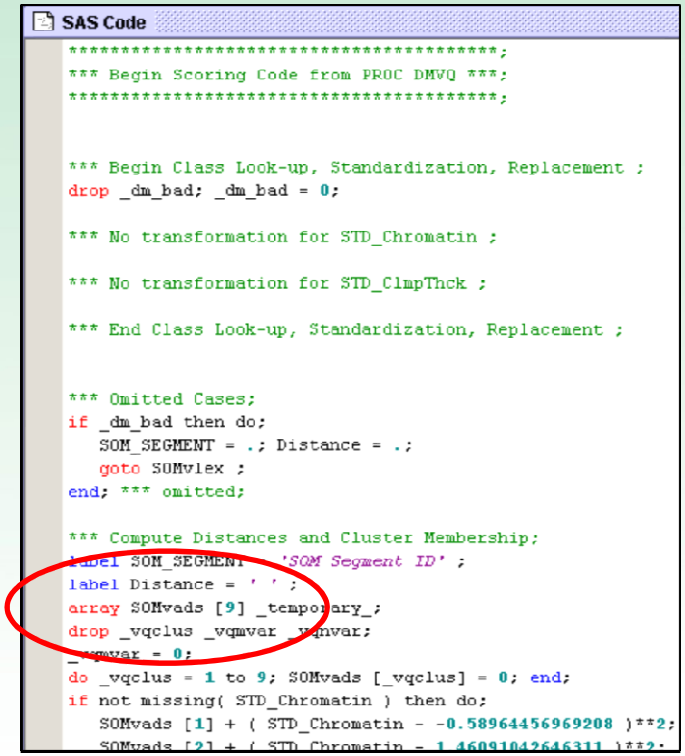
- STEP 5: Paste the SAS Score Code from **STEP 2** inside of the data step.

Approximating Fuzzy with Kohonen/SOM

- STEP 6: Scroll to the top of the code in **STEP 5**. Locate the “distance” array near the top of program.

- STEP 7: Record the name of the array (in this case SOMvads).

“SOMvads” →



Approximating Fuzzy with Kohonen/SOM

•STEP 8: Go to the end of the data step.

• STEP 9: Make the call to the SAS Macro program to calculate fuzzy membership.

• %fuzzyCluster	Name of the macro
• SOMvads	Name of the array variable that holds the distances
• Ufuzzy	Name (given by user) to hold the fuzzy memberships
• SIZE = 9	User tells macro that there are 9 clusters
• M = 2	User tells macro to use an M value of 2

```
else do; SOM_DIMENSION1 = .; SOM_DIMENSION2 = .; SOM_ID = ' '; end;
SOMvlex :;

*****************************************;
*** End Scoring Code from PROC DMVQ ***;
*****************************************;

%fuzzyCluster( SOMvads , Ufuzzy , SIZE=9, M=2 );

run;
```

SOM / KOHONEN NET

Add a call to the macro program

- Multiple calls to the macro are possible, so that the user may calculate different fuzzy memberships for different values of *m*.
- The variable &DIST is a macro variable holding the distance array. It is used for convenience.

```
   _dm8 = left(_dm8); drop _dm8;
   substr( SOM_ID , _vqlen+1 ) = _dm8;
end;
else do; SOM_DIMENSION1 = .; SOM_DIMENSION2 = .; SOM_ID = ' '; end;
SOMvlex :;

*************************************;
*** End Scoring Code from PROC DMVQ ***;
*************************************;

%fuzzyCluster( &DIST. , Ua , SIZE=&ARRAYSIZE., M=1.01 );
%fuzzyCluster( &DIST. , Ub , SIZE=&ARRAYSIZE., M=1.10 );
%fuzzyCluster( &DIST. , Uc , SIZE=&ARRAYSIZE., M=1.25 );
%fuzzyCluster( &DIST. , Ud , SIZE=&ARRAYSIZE., M=1.30 );
%fuzzyCluster( &DIST. , Ue , SIZE=&ARRAYSIZE., M=1.35 );
%fuzzyCluster( &DIST. , Uf , SIZE=&ARRAYSIZE., M=1.40 );
%fuzzyCluster( &DIST. , Ug , SIZE=&ARRAYSIZE., M=2.00 );
```

SAS Enterprise Miner Fuzzy Cluster Approximation

Advantages

- Flexible and can be maintained
- Can have multiple calls to the macro
- Portable (Can be used outside of Enterprise Miner)

Disadvantages

- Non GUI approach takes time to learn
- Requires knowledge of SAS programming and SAS macros

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- **Example showing Fuzzy can improve accuracy**
- Conclusion

Example of Improved Accuracy

- KDD Cup Data Set from 1998. (Parsa and Howes)
  - Cost to send a letter is $0.68
  - Predict which people will donate and therefore improve profit.

- This model will only use a simple 4x4 Kohonen network as a predictor model.
  - WEALTH1 Categorical variable from 0 (lowest) to 9 (highest)
  - TOTAL NUMBER OF GIFTS This is the total number of times a donor gave a gift.
  - TOTAL AMOUNT This is the total amount of money donated by the donor.
  - TIME SINCE LAST DONATION This is the number of months since last donation.

- This is not an optimal solution.

- The purpose is to demonstrate that Fuzzy Membership Approximation can improve accuracy

Example of Improved Accuracy (pg. 2)

PROFILES	1	2	3	4
1	**(5.1 %) -$0.054** Low income and little wealth. They have not given many times and have not given much. They have not given in a long time.	**(6.2 %) $0.031** Average income and above average wealth. They have not given many times and have not given much. They have not given in a long time.	**(5.7 %) -$0.112** Low income and little wealth. Have given frequently in the past and gave given much. Lapse time is about average.	**(3.8 %) $0.085** Lower income and less wealth. Have given frequently in the past and gave given much. They have not given in a long time.
2	**(8.4 %) -$0.082** Low income and little wealth. They have not given many times and have not given much. Lapse time is about average.	**(6.3 %) $0.139** Very high income but have little wealth. They have not given many times and have not given much. Lapse time is about average.	**(7.1 %) $0.132** Average income and little wealth. Have given frequently in the past and gave given much. Lapse time is about average.	**(5.4 %) $0.344** High income and high wealth. They have not given many times and have not given much. They have not given in a long time.
3	**(7.4 %) -$0.062** Low income and high wealth. They have not given many times and have not given much. Lapse time is about average	**(2.7 %) $0.049** Low income and low wealth. Have given an extremely large number of times and an extremely large sum of money. Lapse time is average.	**(10.4 %) -$0.056** Low income and low wealth. They have not given many times and have not given much. Lapse time is average.	**(12.3 %) $0.120** Average income and high wealth. They have not given many times and have not given much. Lapse time is average.
4	**(0.3 %) $2.790** Average income and wealth. Have given significant amounts of money in the past and have given frequently. Have given more recently than the average.	**(3. 4 %) $0.599** Average income and wealth. Have given frequently in the past and gave much. Have given much more recently than the average.	**(7.4 %) $0.373** High income and high wealth. Have given frequently in the past and gave much. Lapse time is average.	**(7.9 %) $0.411** High income and high wealth. They have not given many times and have not given much. Lapse time is average.

Example of Improved Accuracy (pg. 3)

- Each cluster has a profile description and average profit.
- For HARD clusters, simply do not send to the clusters with negative average profit.
- For FUZZY clusters, use a weighted average approach and don't send to negative average profit.
- Example, assume a person was
  - 60% in (3,3)
  - 40% in (3,4)

Example of Improved Accuracy (pg. 4)

PROFILES	1	2	3	4
1	**(5.1 %) -$0.054** Low income and little wealth. They have not given many times and have not given much. They have not given in a long time.	**(6.2 %) $0.031** Average income and above average wealth. They have not given many times and have not given much. They have not given in a long time.	**(5.7 %) -$0.112** Low income and little wealth. Have given frequently in the past and gave given much. Lapse time is about average.	**(3.8 %) $0.085** Lower income and less wealth. Have given frequently in the past and gave given much. They have not given in a long time.
2	**(8.4 %) -$0.082** Low income and little wealth. They have not given many times and have not given much. Lapse time is about average.	**(6.3 %) $0.139** Very high income but have little wealth. They have not given many times and have not given much. Lapse time is about average.	**(7.1 %) $0.132** Average income and little wealth. Have given frequently in the past and gave given much. Lapse time is about average.	**(5.4 %) $0.344** High income and high wealth. They have not given many times and have not given much. They have not given in a long time.
3	**(7.4 %) -$0.062** Low income and high wealth. They have not given many times and have not given much. Lapse time is about average	**(2.7 %) $0.049** Low income and low wealth. Have given an extremely large number of times and an extremely large sum of money. Lapse time is average.	**(10.4 %) -$0.056** Low income and low wealth. They have not given many times and have not given much. Lapse time is average.	**(12.3 %) $0.120** Average income and high wealth. They have not given many times and have not given much. Lapse time is average.
4	**(0.3 %) $2.790** Average income and wealth. Have given significant amounts of money in the past and have given frequently. Have given more recently than the average.	**(3. 4 %) $0.599** Average income and wealth. Have given frequently in the past and gave much. Have given much more recently than the average.	**(7.4 %) $0.373** High income and high wealth. Have given frequently in the past and gave much. Lapse time is average.	**(7.9 %) $0.411** High income and high wealth. They have not given many times and have not given much. Lapse time is average.

60 % **40%**

Example of Improved Accuracy (pg. 5)

- Example, assume a person was
  - 60% in (3,3) - $0.056 loses money
  - 40% in (3,4) $0.120 earns money

- In HARD CLUSTERING, we would NOT send the person a letter.
  - Person is Closer to Cluster that LOSES MONEY

Example of Improved Accuracy (pg. 6)

- Example, assume a person was
  - 60% in (3,3)
  - 40% in (3,4)
- Profit for Each Cluster
  - Average profit for (3,3) is -$0.056
  - Average profit for (3,4) is $0.120
- Profit Calculations

Profit = 0.60*(-$0.056) + 0.40* $0.120
= -$0.0336 + $0.048
= $ 0.0144

This is POSITIVE so send this person a request

Example of Improved Accuracy

TRAINING DATA			
Model	**% Customers Solicited**	**Average Profit**	**Total Profit**
Send To All	100 %	$0.12	$5804
Hard Cluster	63 %	$0.24	$7067
Fuzzy M=1.01	63 %	$0.24	$7060
Fuzzy M=1.10	64 %	$0.23	$7026
Fuzzy M=1.25	65 %	$0.23	$6998
Fuzzy M=1.30	66 %	$0.23	$7264
Fuzzy M=1.35	67 %	$0.23	$7167
Fuzzy M=1.40	68 %	$0.22	$7080
Fuzzy M=2.00	97%	$0.13	$5815
Fuzzy M=3.00	100%	$0.12	$5804

Example of Improved Accuracy

TEST DATA			
Model	**% Customers Solicited**	**Average Profit**	**Total Profit**
Send To All	100 %	$0.10	$4985
Hard Cluster	62 %	$0.19	$5509
Fuzzy M=1.01	62 %	$0.19	$5556
Fuzzy M=1.10	63 %	$0.19	$5676
Fuzzy M=1.25	64 %	$0.19	$5777
Fuzzy M=1.30	65 %	$0.19	$5799
Fuzzy M=1.35	66 %	$0.18	$5763
Fuzzy M=1.40	67 %	$0.17	$5593
Fuzzy M=2.00	97%	$0.11	$4979
Fuzzy M=3.00	100 %	$0.10	$4985

Example of Improved Accuracy

- Both “hard” and “fuzzy” were better than “send to everybody” on training data. Fuzzy was better.

Hard Cluster	**Avg. Profit = + $0.12**	**Total Profit = + $1,263**
Fuzzy Cluster	**Avg. Profit = + $0.11**	**Total Profit = + $1,460**

- Both “hard” and “fuzzy” were better than “send to everybody” on test data. Fuzzy was better.

Hard Cluster	**Avg. Profit = + $0.09**	**Total Profit = + $524**
Fuzzy Cluster	**Avg. Profit = + $0.09**	**Total Profit = + $814**

55% Improvement over HARD CLUSTERS

… by simply calling a SAS MACRO

- What is Clustering
- How is Data Clustered
  - K Means Clustering
  - Kohonen/SOM Clustering
  - Advantages and Disadvantages of K-Means and Kohonen/SOM
- Difficulties with Hard Clusters

- Fuzzy Clustering
  - Fuzzy Logic
  - Fuzzy C-Means Algorithm
  - Fuzzy Membership Function
  - Example Of Fuzzy Membership
  - Advantages and Disadvantages of Fuzzy Clustering

- Fuzzy Cluster Approximation (Hard/Fuzzy Hybrid)
- Fuzzy Cluster Approximation in Enterprise Miner
  - Fuzzy Membership Macro
  - K-Means
  - Kohonen / SOM Clustering
- Example showing Fuzzy can improve accuracy
- Conclusion

CONCLUSIONS

- Fuzzy Clusters can be approximated in SAS Enterprise Miner
- Requires only a small amount of extra work for user
- Can have significant impact on predictive modeling

CONTACT

- If you wish to receive a copy of:
  - Slide Deck
  - SAS Macro

- Contact:

don.wedding@sas.com