

# Fuzzy Logic and Clustering - Harrison Kiang

## Introduction

Many K-partitioning clustering algorithms such as K-Medians, K-Modes, K-Medoids, etc. depend on object cluster assignment and centroid update. That is, to find the cluster assignments of  $n$  points  $X = \{x_1, \dots, x_n\}$ , to  $k$  clusters  $C = \{c_1, \dots, c_k\}$ , each iteration has a set of common steps:

1. Assign each point  $x \in X$  to the nearest cluster  $c \in C$
2. For every cluster  $c \in C$ , calculate the new position of the respective centroid.

In the cluster analysis part of CS 412: Introduction to Data Mining by professor Jiawei Han from the University of Illinois at Urbana-Champaign, there is a lecture in the “Partitioning-Based clustering Methods” section discussing “The K-Medians and K-Modes Clustering Methods”. Here is the link: [3.5 The K-Medians and K-Modes Clustering Methods](#) (only accessible to those enrolled in the course). An alternative for K-Modes, fuzzy K-Modes, was mentioned: “A **fuzzy K-Modes** method is proposed to calculate a **fuzzy cluster membership value** for each object to each cluster”. Unfortunately, this important concept of fuzzy clustering is not covered in detail in the lecture.

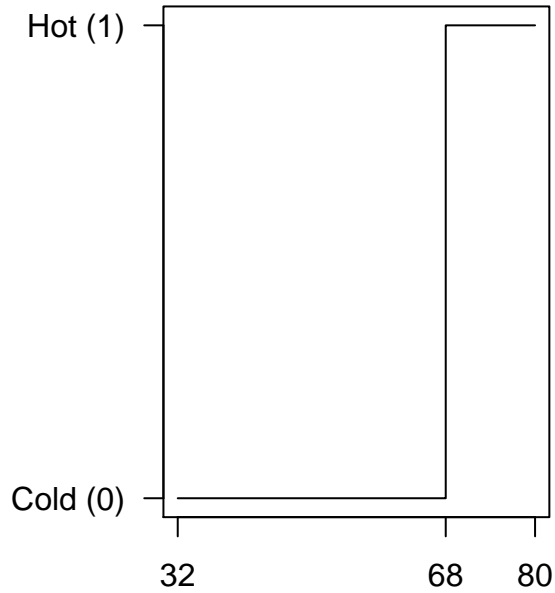
The purpose of this paper is to serve as an overview as to what fuzzy clustering is, and include a comparison of fuzzy clustering and hard clustering. An algorithm employing fuzzy clustering called the Fuzzy c-Means Algorithm will also be discussed.

## Fuzzy Logic

If you ask someone to describe how the temperature feels outside, they may either say it’s hot or it’s cold. Here, there is no precise mathematical definition for “hot” or “cold”. You could also ask someone to describe a person’s height. They may say that a person is tall or short. Here, there is no precise mathematical definition for “tall” or “short”.

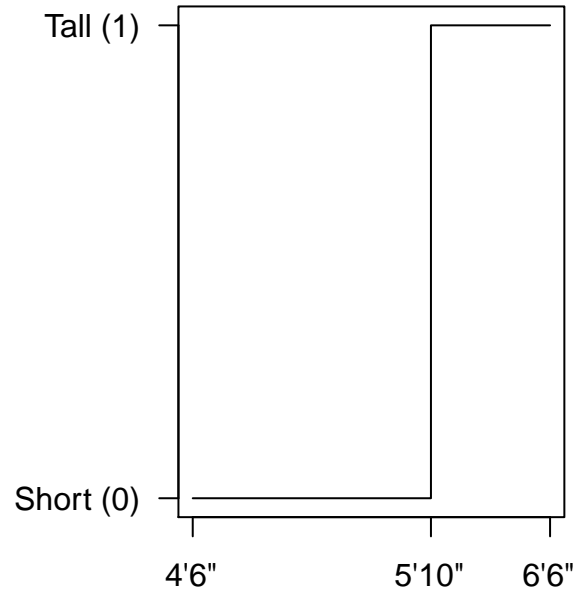
In both cases, it’s possible to define a crisp boundary between what is defined as hot or cold, and between what is defined as tall or short. One could say that any temperature less than 68°F is cold, and any temperature greater than or equal to 68°F is hot. One could also say that any height less than 5 feet 10 inches is short, and any height greater than or equal to 5 feet 10 inches is tall. Here, 68 ° and 5 feet 10 inches are the crisp boundaries for temperature and height, respectively.

### Hot and Cold Crisp Boundary



Temperature in Degrees Farenheit

### Tall and Short Crisp Boundary



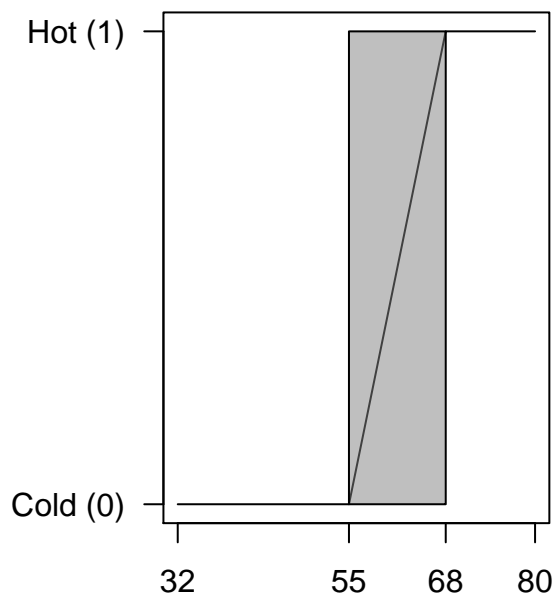
Height in Feet and Inches

Temperature and height are both interperable fields in that some people might interpret a given temperature or height differently. In other words, some people might classify a person as tall, and others might classify that same person as short. Likewise, some people might consider a given temperature as cold, and others might classify that same temperature as hot.

There are certain heights in which the vast majority of people would agree to be either “short” or “tall” (likewise for “short” and “tall”). Let’s say that everyone agrees that people 5 feet and 2 inches and below in height are short, and people 5 feet and 10 inches and above in height are tall. Let’s say that everyone agrees that any temperature less than or equal to 55°F is cold, and any temperature greater than or equal to 68°F.

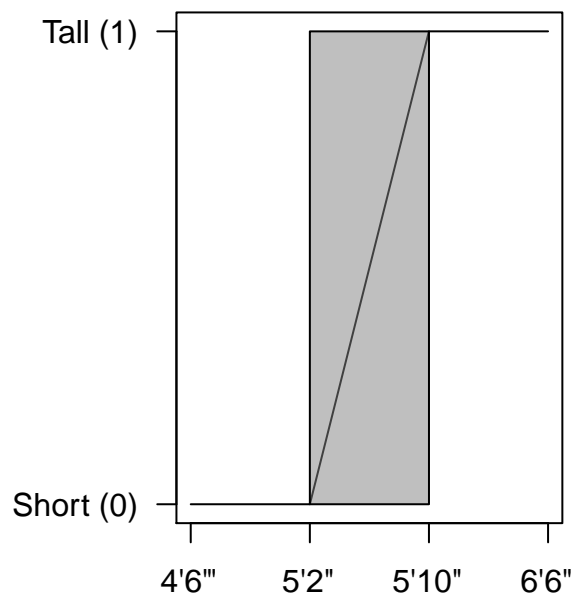
The area of disagreement, or the “grey area”, is where the value of the sample is uncertain. That is, for samples  $s$  within this “grey area”,  $s \in [0, 1]$ . A value of 0 or 1 means that everyone agrees on the attribute value. A value in between 0 and 1 indicates a degree of uncertainty. Such a distribution for both temperature and height with “grey areas” is shown below:

## Hot and Cold Crisp Boundary



Temperature in Degrees Farenheit

## Tall and Short Crisp Boundary



Height in Feet and Inches

From here, you can make decision rules based on the values in graphs like the one above to drive decisions in the real world. For example, if it's cold and windy, one may as a result decide to layer up and vice versa. If one is in the midst of a group of taking a photo and they are of short height, they may appear best in the photo if they place themselves in the front row. These are just a few examples.

Here is an excellent video with greater detail on on fuzzy logic with a number of real world examples:

<https://youtu.be/r804UF8Ia4c>

## Fuzzy Clustering vs Hard Clustering

Fuzzy clustering, or soft clustering, is where data points have the potential to belong to multiple clusters. Unlike clustering, or hard clustering, where points belong to exactly one cluster, points have the ability to be members of multiple clusters.

The membership between points and clusters is stored in a  $n \times k$  matrix  $U$  where  $u_{ij} \in [0, 1]$  is the degree to which  $x_i$  belongs to cluster  $c_j$ . Note that  $\sum_{j=1}^k u_{ij} = 1$  for every row entry  $u_i$  in  $U$ . This means that  $\sum_{i=1}^n \sum_{j=1}^k u_{ij} = n$ . That is, the degree to which points belong to all clusters sum to 1, and therefore the degree to which all points belong to all clusters is equal to the number of points. A sample membership matrix is shown below:

$$\mathbf{U}_{\text{fuzzy}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.5 & 0.5 \\ 0.4 & 0 & 0.6 \end{bmatrix}$$

Here, point  $x_1$  is only a member of cluster  $c_3$ , point  $x_2$  is equally a member of both clusters  $c_2$  and  $c_3$ , and point  $x_3$  is a stronger member of  $c_3$  compared to  $c_1$ .

In hard clustering, such a membership matrix might look like this:

$$\mathbf{U}_{\text{hard}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here, point  $x_1$  is only a member of cluster  $c_3$ , point  $x_2$  is only a member of cluster  $c_2$ , and point  $x_3$  is only a member of  $c_3$ . Note in both variants that for every row entry  $u_i$  in  $U$ ,  $\sum_{j=1}^k u_{ij} = 1$ , and therefore  $\sum_{i=1}^n \sum_{j=1}^k u_{ij} = n$ .

## Fuzzy C-Means Clustering (FCM) Algorithm

Fuzzy C-Means Clustering is very similar to the K-Means algorithm. It takes the following as inputs:

1. The number of clusters  $k$
2. The points to cluster  $X$
3. Initial membership entries  $u_{ij_0} \in U$
4. A termination criterion  $0 < \epsilon < 1$

The basic outline of the Fuzzy C-Means Algorithm is much like K-Means. Fuzzy C-Means is an iterative function that assigns clusters to points, but in varying degree. Fuzzy-C means also recalculates the location of clusters taking the entire matrix  $U$  into account. That is, the cluster locations are a function of the degree of membership for the clusters, respectively.

The objective function for Fuzzy C-Means is as follows:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty$$

$m$  is the “fuzzifier”. A large  $m$  results in smaller membership values  $u_{ij}$  and therefore fuzzier clusters. If  $m = 1$ , the memberships converge to 0 or 1, resulting in crisp partitioning.  $m$  is typically set to 2 in the absence of domain knowledge.

$u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $c_j$ . These are entries in the matrix  $U$  as stated above.

$\| * \|$  is any norm expressing similarity between any measured data and the cluster. This can be any distance function, therefore allowing for flexibility in the algorithm. Common distance functions include variations of the Minkowski Distance ( $\sum_{i=1}^n |x_i - y_i|^p$ )<sup>1/p</sup> Manhattan distance, or  $L_1$  norm ( $p = 1$ ), Euclidean distance, or  $L_2$  norm ( $p = 2$ ), or the supremum distance ( $p \rightarrow \infty$ ).

Each iteration involves updating entries  $u_{ij} \in U$  and the cluster centers  $c_j \in C$  as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^N u_{ij} x_i}{\sum_{i=1}^N u_{ij}^m}$$

Note that like K-Means, Fuzzy C-Means is only able to find a local minimum, and clustering results will differ depending on the initial entries  $u_{ij}$  in  $U$ . For Fuzzy C-Means, the saddle or local minimum of  $J_m$  is found.

The local minimum will be found when the following condition is met:

$$\max_{ij} \{|u_{ij}^{(k+1)} - u_{ij}^{(k)}|\} < \epsilon$$

where  $\epsilon$  is the termination criterion as inputted above.

Here is a link to an Interactive Demo for Fuzzy C-Means:

[https://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/AppletFCM.html](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/AppletFCM.html)

## Resources

[https://en.wikipedia.org/wiki/Fuzzy\\_clustering#Comparison\\_to\\_Hard\\_Clustering](https://en.wikipedia.org/wiki/Fuzzy_clustering#Comparison_to_Hard_Clustering)

<https://www.youtube.com/watch?v=P8wY6mi1vV8>

<https://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html>

[https://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/cmeans.html](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html)

<https://homes.di.unimi.it/valenti/SlideCorsi/Bioinformatica05/Fuzzy-Clustering-lecture-Babuska.pdf>