

Executive Summary (1 page or less)

Fill in by October 24th and update

- 1) Briefly describe your proposed idea (paragraph or two, this can be an updated version of your proposal). What's the concept you'll be explaining?

I will explain the idea of k-means clustering (https://en.wikipedia.org/wiki/K-means_clustering). k-means clustering is an iterative algorithm applied in cluster analysis. K-means clustering aims to partition a set of points distributed within a space into several clusters, whose number determined by the user, according to their locations. Its fundamental procedures are, first randomly initiate some label points on the space, and assign all the points closet to each label as one group, then move the labels to the center of the group points, and reassign all the points into new group according to their distance to the new labels. This procedure is repeated until a stability is achieved, that is labels are all in the center of each group, and no points can be assigned into a new group. K-means is very fundamental but practical, and is widely used in machine learning and data analysis.

- 2) What level are you trying to teach for (high school? college? graduate school?). You should probably not be aiming at elementary school level, but if you have a great idea that's ok.

This project aims at high school level.

Learning Objectives (1 page or less)

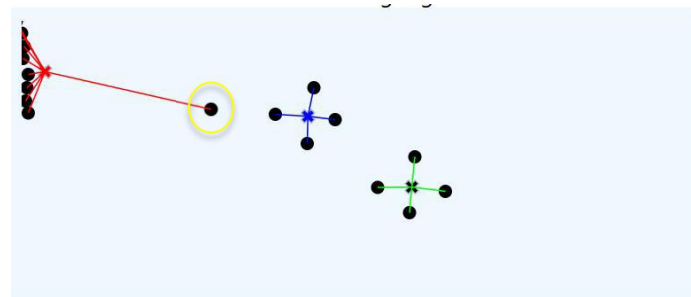
Fill in by October 24th and update

3-5 learning objectives for your project.

1. The viewer will be able to describe the overall procedures of k-means algorithm
2. the viewer will be able to predict how the group update stage works in the k-means algorithms.
3. The viewer will be able to handle special case or special case in k-means algorithm

Create an Exam

1. multiple choice: the k-algorithm takes place in which of the following order?
(a) update group - move center - update group - move center - ...
(b) move center - update group - move center - update group - ...
(c) Update group - move center - move center - update group -...
2. Multiple choice: in the following graph, black dots are clustering objectives, and cross are group centers. Which group will the yellow circled black dot be clustered into in next stage?

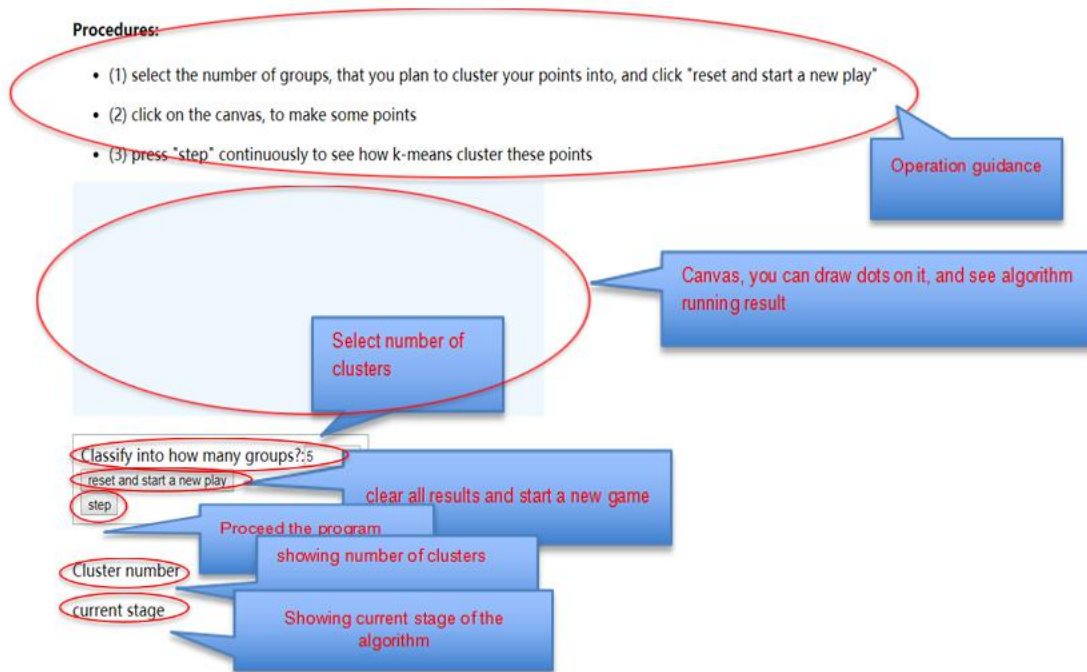


- (a) red cross group
 - (b) Blue cross group
 - (c) Black cross group
3. Open answer: please draw a case on the paper, that none of the points are clustered into a certain group, and explain why.

Your sketch/screenshots (as many pages as you need)

Fill in by October 24th and update

Introduction of the interface



Step 1: setting and start

Procedures:

- (1) select the number of groups, that you plan to cluster your points into, and click "reset and start a new play"
- (2) click on the canvas, to make some points
- (3) press "step" continuously to see how k-means cluster these points



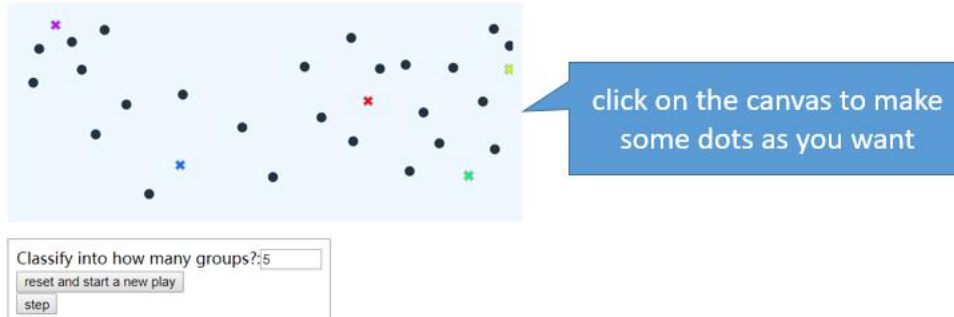
cluster number: 5

current stage

Step 2: create dots

Procedures:

- (1) select the number of groups, that you plan to cluster your points into, and click "reset and start a new play"
- (2) click on the canvas, to make some points
- (3) press "step" continuously to see how k-means cluster these points



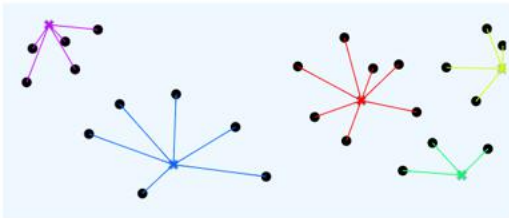
cluster number: 5

current stage

Step 3: proceed the program

Procedures:

- (1) select the number of groups, that you plan to cluster your points into, and click "reset and start a new play"
- (2) click on the canvas, to make some points
- (3) press "step" continuously to see how k-means cluster these points



Classify into how many groups?:5
reset and start a new play
step

Press the step, for the program to undergo each step of the k-means

cluster number: 5

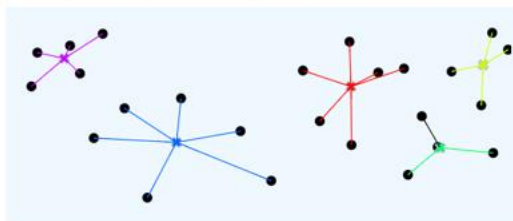
current stage: update center of each group

Current step is shown here

Step 4: complete the process, and restart

Procedures:

- (1) select the number of groups, that you plan to cluster your points into, and click "reset and start a new play"
- (2) click on the canvas, to make some points
- (3) press "step" continuously to see how k-means cluster these points



If the dots and centers remain the same for several stages, its means the process is over

Classify into how many groups?:5
reset and start a new play
step

You can start a new game if you want

cluster number: 5

current stage: update center of each group

Why will your vis help? (2 page or less)

Fill in by October 24th and update

K-means clustering is traditionally presented with complex math calculation, entailing mean and distance calculation. It is an iterative process, which lead to more difficulty in understanding the algorithm structure. All these features can be problematic for people with relatively weak mathematical basis to capture the core idea of k-means clustering.

this project, however, applies visualization to demonstrate the process of k-means. Although it is not as abundant in details as mathematical expressions, it is easier to acknowledge for the wider audience who are not so proficient in Math. Moreover, the mathematical details are not really vital in understanding the general concepts of k-means, so this visualization is obviously a more desirable guidance for people interested in k-means algorithm.

This visualization mainly use shape and color as encoding features. The objects to be clustered is black dots, and the center of each group is cross. It also uses color to represent each different group, and connect all dots in one group with line. In that way, the encoding is very clear, and easy to understand. It is also very pre-attentive, since you can easily tell which dot is in which group with their color and connecting lines.

The interactive part is the key to educating the viewers. First, the viewer can freely create all the dots to be clustered. They can freely decide dots' locations and distributions as their wish, and see how this will affect the final clustering results. Second is the parameters controlling. The viewer can freely determine the number of clusters, so they can see: even with same number and distribution of dots, how will the cluster result change if the cluster number is different? They also have choice to renew the initial cluster center. According to the algorithms, initial centers are all randomly generated, so they can see how will different initialization will impact the clustering. Third, there is a "step" button, so viewers can control the program running step by step, and clearly see what happens within each step.

I think the point to be improved is that, maybe more clustering algorithms can be added in for users to choose, so they can better compare between different ones and acquire a deeper learning of clustering.

Archive of Experiments (as many pages as you need)

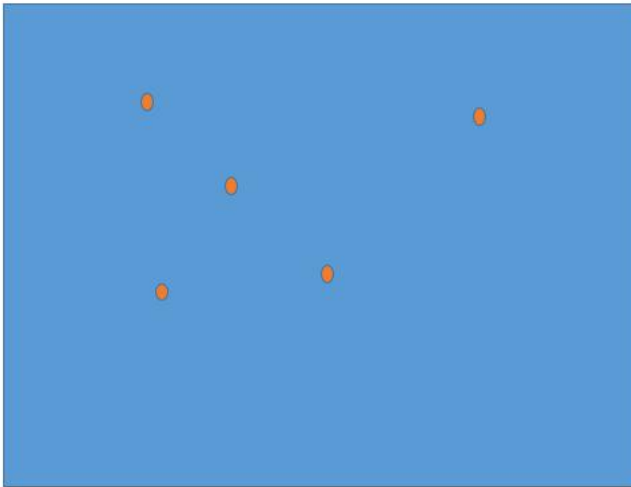
Start by October 24th, and updated as you go

Clean result

Start (Generate
centers)

step

This is my original interface. Later, I realize that when people clean result, they will also start a new game, and it is inconvenient to have so many buttons, so I combine them and make “reset and start a new game” button.



Number of nodes

Number of clusters

The second version is like above. The users can only decide number of dots, but they can not decide their location, which is determined by a random generator. I think if viewer can decide the location on their own, they can learn more from the visualization, so I toss this.

Inspiration/Competitive Analysis

Start by October 24th, and updated as you go

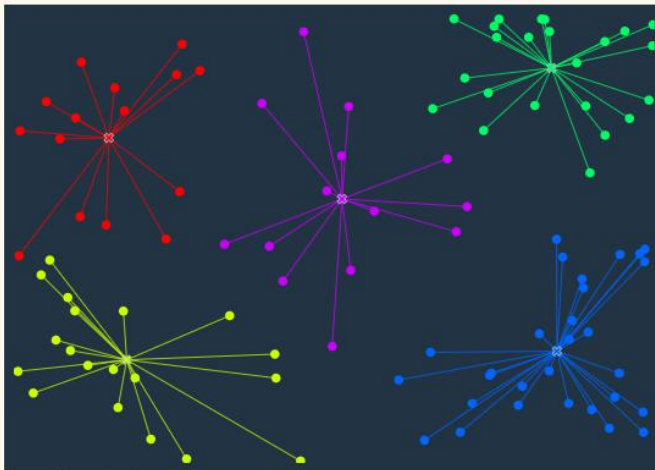


07 Nov 2013 JavaScript

Visualizing K-Means algorithm with D3.js

Tweet Like 84

The K-Means algorithm is a popular and simple clustering algorithm. This visualization shows you how it works.

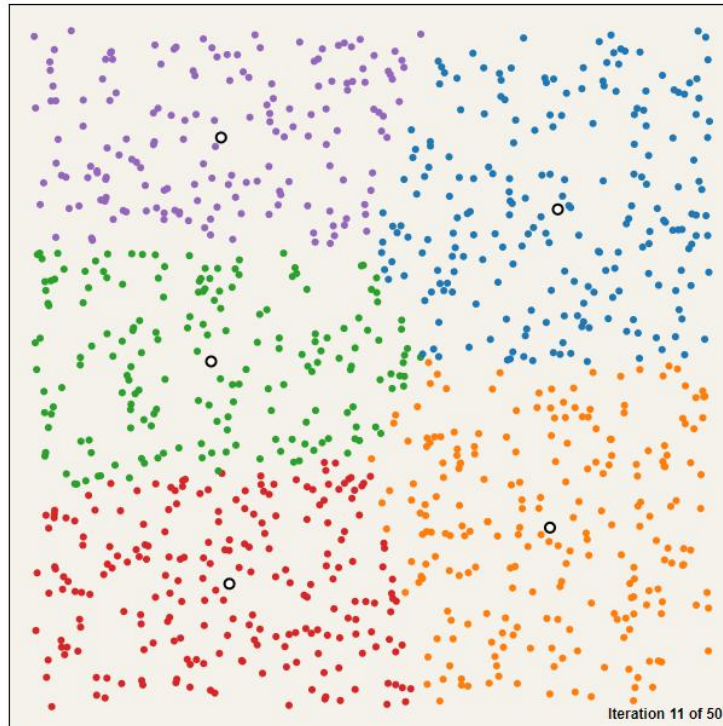


Step Restart

N (the number of node)	<input type="text" value="100"/>
K (the number of cluster)	<input type="text" value="5"/>
<input type="button" value="New"/>	

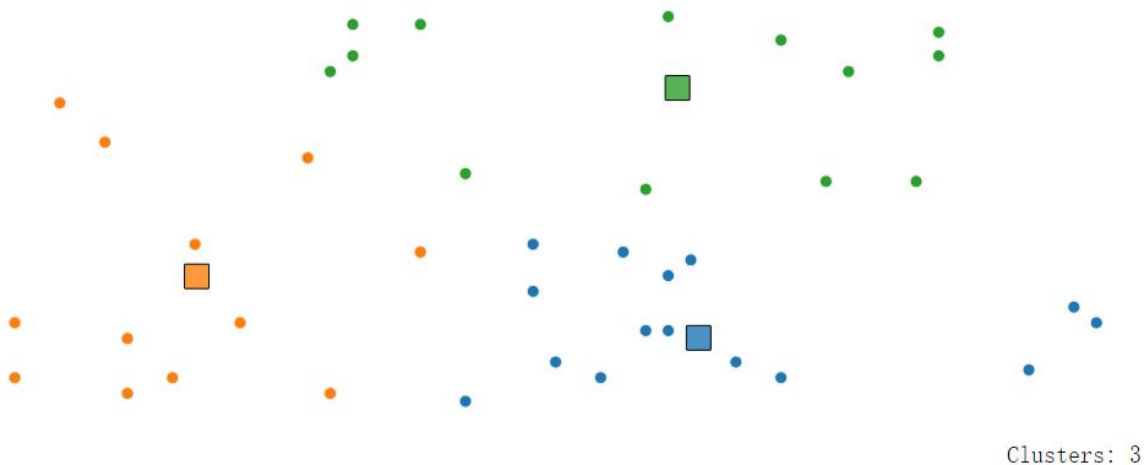
This is a visualization from nitoyon (<http://tech.nitoyon.com/en/blog/2013/11/07/k-means/>). the user can decide number of dots and clusters, but not the dot location, which I realize in my program.

K-means clustering



Here is a program from Wolf Dynamic (<https://bl.ocks.org/wolfDynamics/8682a6ff548e820d4acb5cd0e87ca603>), it shows the process of k-means, but has totally no explanation and interactive elements, so it is hard for users to understand what is happening.

Assign and Update (3).



This is from mayblue9 (<https://bl.ocks.org/mayblue9/fb21145b21025a411150c158df979221>). It shows the process of k-means, with number of clustering shown on the right down corner, and stage on the top, and will repeat stage for several different dots and centers. But it also have the problem of no interacting. The users can decide nothing, so they can not learn by make comparisons between different parameters and conditions. The program keep running automatically, so it is inconvenient if users want to pause and observe a certain stage more carefully.

Peer-Robin Notes (If you get interesting feedback you want to share)

1. There should be explanation to the basic ideas of k-means
2. there should be guidance for users to know how use the interface.
3. can use different colors and shapes for users to more clearly see the information