

Merritt Burch

Buckler Lab Presentation

2019-04-25

George Harrison Shull (1872-1954)



PhD from University of Chicago in 1904

Described heterosis and coined term in maize by crossing stable inbreds together to create vigorous hybrids (1908), estimated to have increased corn yields by 20% (+1.8 billion bushels = \$2 billion dollars); founder of 'Genetics' journal, worked on many different crops (primrose, sunflower, tomato, poppy, tobacco, etc.)

Botanical expert to the Bureau of Plant Industry; Botanical investigator at Carnegie Institution at Cold Spring Harbor; Princeton University (started in 1915)

1948 Public Welfare Medal from the National Academy of Sciences

Never completed a full-year of grade school. Went on to be a country school teacher in Ohio for 5 years, taught 7 grades/day, tried to become a principal in 1901

K-means clustering of multi-dimensional points

These are special cases of the general definition: given two n -dimensional points

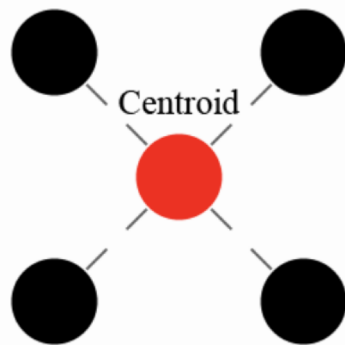
$\mathbf{p} = (p_1, p_2, \dots, p_n)$ and $\mathbf{q} = (q_1, q_2, \dots, q_n)$ the Euclidean distance between them is:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

- Centroids are the mean of cluster points
- Points belong to cluster whose mean/centroid closest to

Steps to k-means clustering

- 1) Pick k centroids (randomly or through seeds)
- 2) Partition the data into centroids
- 3) Recompute centroid means with points
- 4) Repeat 2 and 3 until convergence (means stop changing)



Going to attempt to cast code/example to screen