

Clustering

Nathaniel Lewis

How it works

Read in Historic Data

Generate Centroids randomly

Assign data points to Centroids

Average values of data points and adjust Centroids

Repeat 3 & 4 until no data points are reassigned

Read in new data and predict outcome based on closest centroid

K-Modes

Derivative of k-means

Works with Nominal data

Uses number of different answers to determine distance

Centroid values are adjusted to the mode of data points assigned to it

Issues I had

<Template ItemType>

Object Linking

Segmentation Faults

Summary

K-Means finds clusters in numeric data

K-Modes finds clusters in nominal data

Clusters are used in predictions

Programming is hard.

Citations

Coates, A., & Ng, A. Y. (1970, January 01). Learning Feature Representations with

K-Means. Retrieved November 26, 2017, from https://link.springer.com/chapter/10.1007/978-3-642-35289-8_30

[doi 10.1007/s11004-010-9276-7](https://doi.org/10.1007/s11004-010-9276-7)

K-modes. (2014, September 14). Retrieved November 26, 2017, from

<https://shapeofdata.wordpress.com/2014/03/04/k-modes/>

["Least squares quantization in PCM"](#)

[IEEE](#)

[Transactions on Information Theory](#)

[doi 10.1109/TIT.1982.1056489](#)

Citations 2

[Some Methods for classification and Analysis of Multivariate](#)

[Observations](#)

[MR 0214227](#) [Zbl 0214.46201](#)

[Steinhaus, H.](#)

[MR 0090073](#) [Zbl 0079.16403](#)

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (n.d.). *Constrained K-means*

Clustering with Background Knowledge (2001 ed., Proceedings of the Eighteenth International Conference on

Machine Learning, pp. 577-584, Rep.).