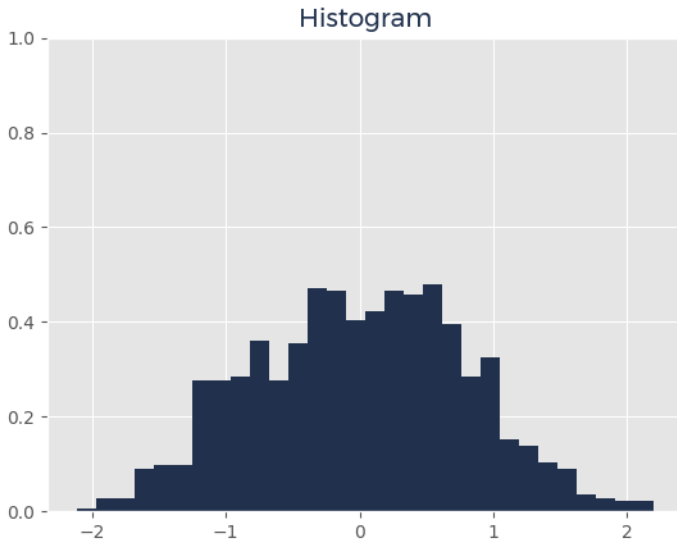


Unsupervised Learning - Density Estimation

Creating our own truth

Histograms



Histogram Bin Width

Scott's rule

$$w = 3.49\sigma n^{\frac{-1}{3}},$$

where σ is the standard deviation of the data and n is the number of samples.

Histogram Bin Width

Freedman and Diaconis' rule

$$w = 2dn^{\frac{-1}{3}},$$

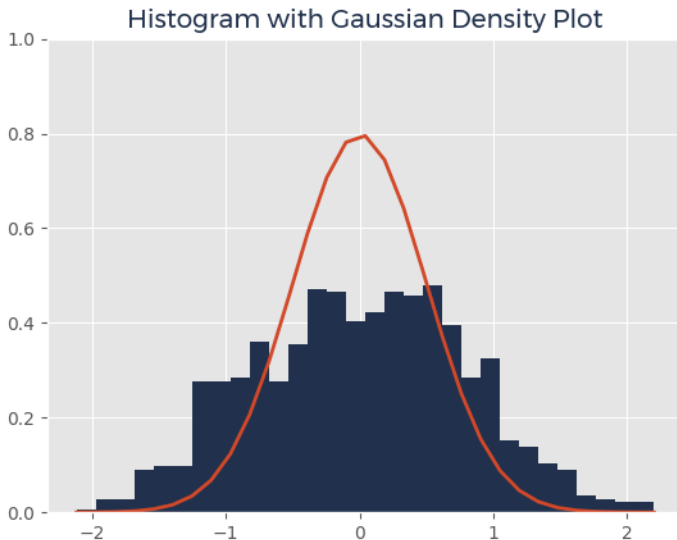
where d is the interquartile range (difference between 75th percentile and 25th percentile), and n is the number of samples.

Histogram Bin Width

- Knuth's rule
- Bayesian block representations

Demo

Kernel Density Estimation



Kernel Density Estimation

The kernel density estimator of an unknown density f is,

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

where n is the number of samples, h is the bandwidth which is a smoothing parameter, and $K()$ is a kernel applied to the data.

$$K_h(x_i) = \frac{1}{h} K\left(\frac{x_i}{h}\right).$$

Gaussian

$$K_h(x) \propto \exp \frac{-x^2}{2h^2},$$

Uniform / Tophat

$$K_h(x) \propto 1 \text{ if } x < h,$$

Exponential

$$K_h(x) \propto \exp \frac{-x}{h},$$

Linear

$$K_h(x) \propto 1 - \frac{x}{h},$$

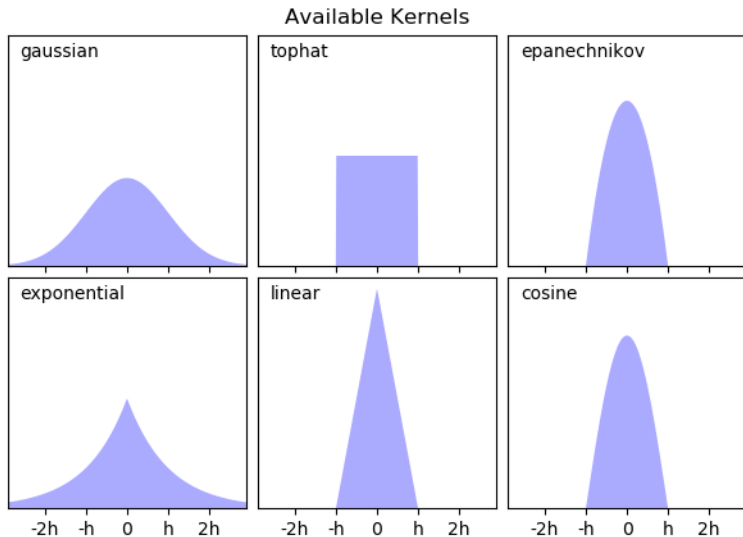
Epanechnikov

$$K_h(x) \propto 1 - \frac{x^2}{h^2},$$

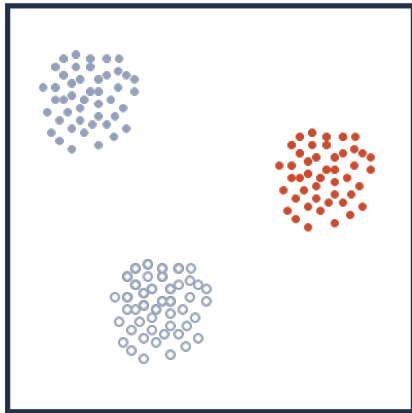
Cosine

$$K_h(x) \propto \cos \frac{\pi x}{2h}.$$

Kernels



Demo



Clustering

Density-Based Clustering Methods

- Gaussian Mixture Models
- MeanShift
- DBSCAN

Gaussian Mixture Models

Clustering by assuming each cluster follows a Gaussian distribution.

$$\mathbf{x}_{i,t+1} = \mathbf{x}_{i,t} + \mathbf{m}_{i,t} \quad (1)$$

$$\mathbf{m}_i = \frac{\sum_{j \in N(\mathbf{x}_i)} K(\mathbf{x}_i - \mathbf{x}_j) \mathbf{x}_j}{\sum_{j \in N(\mathbf{x}_i)} K(\mathbf{x}_i - \mathbf{x}_j)} \quad (2)$$

where the set $N(\mathbf{x}_i)$ is the set of points within a specified distance of \mathbf{x}_i called its *neighborhood*, and K is a kernel.

Clustering by assigning points a status of core or border.

A core point is one that has a minimum number of points within a particular distance.

All other points are border points.

A cluster is defined as a set of connected core points.

Minimum number of points and the distance to check for in order to make a point a core point are hyperparameters.

Clusters do not have to form convex shapes.

Cluster number is **not** a hyperparameter.

Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.

Some content in these slides are obtained from external sources and may be copyright sensitive. Copyright and all rights therein are retained by the respective authors or by other copyright holders. Distributing or reposting the whole or part of these slides not for academic use is HIGHLY prohibited, unless explicit permission from all copyright holders is granted.