Depth Quantile Functions in Unsupervised Learning

Guy Thampakkul
Pomona College

February 3, 2023

Roadmap

Outline

- Machine Learning
  - Unsupervised Supervised Learning
  - Unsupervised Learning with Supervision
  - Interpretability and Intervention
  - Mainstream Unsupervised Learning Techniques
- Depth Quantile Functions (Chandler, Polonik 2021)
  - Statistical Depth
  - Tukey Depth (1974)
  - Depth Quantile Functions
  - Examples
  - Clustering with DQFs
- Future Directions

Machine Learning

Machine Learning

Machine learning (ML) is a sub-field of Artificial Intelligence (AI) devoted to building models that uses data to improve performance on some set of tasks.

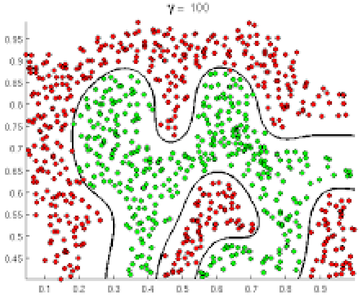
Examples of Machine Learning

- Search Engine Suggestions
- Facial Recognition
- Computer Vision
- Classification
- Clustering

Supervised and Unsupervised Learning

Supervised Learning

Learns functions from data of input-output pairs.
Example: Classification



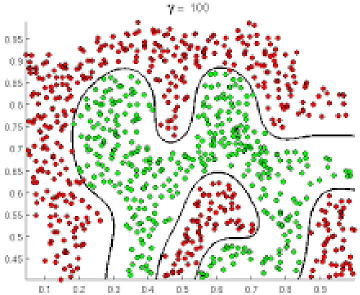
Supervised and Unsupervised Learning

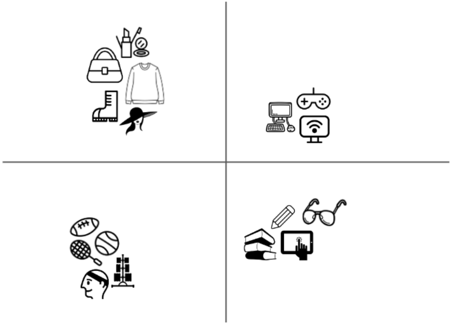
Supervised Learning

Learns functions from data of input-output pairs.
Example: Classification

Unsupervised Learning

Learn patterns from unlabeled data.
Example: Anomaly Detection, Clustering





Some quotes and definitions from *towardsdatascience.com*

Some quotes and definitions from *towardsdatascience.com*

Unsupervised Learning: Clustering

The objective of clustering is to find different groups within the elements in the data. To do so, clustering algorithms find the **structure** in the data so that elements of the same cluster (or group) are **more similar** to each other than to those from different clusters.

Some quotes and definitions from *towardsdatascience.com*

Unsupervised Learning: Clustering

The objective of clustering is to find different groups within the elements in the data. To do so, clustering algorithms find the **structure** in the data so that elements of the same cluster (or group) are **more similar** to each other than to those from different clusters.

Unsupervised Learning: Anomaly Detection

Anomaly detection is the process of identifying **unexpected items** or events in data sets which **differ** from **the norm**.

Unsupervised Learning

Some quotes and definitions from *towardsdatascience.com*

Unsupervised Learning: Clustering

The objective of clustering is to find different groups within the elements in the data. To do so, clustering algorithms find the **structure** in the data so that elements of the same cluster (or group) are **more similar** to each other than to those from different clusters.

Unsupervised Learning: Anomaly Detection

Anomaly detection is the process of identifying **unexpected items** or events in data sets which **differ** from **the norm**.

Problem: Does Unsupervised Learning not need supervision?

Black Box Machine Learning

Black Box

A complex piece of equipment where the contents are mysterious to the user.

Black Box

A complex piece of equipment where the contents are mysterious to the user.

Black Box Machine Learning

Machine learning models that give you a result or reach a decision without explaining or showing how they did so.

Black Box Machine Learning

Black Box

A complex piece of equipment where the contents are mysterious to the user.

Black Box Machine Learning

Machine learning models that give you a result or reach a decision without explaining or showing how they did so.

Neural Networks

- Extremely powerful machine learning technique
- Little to no information of how a trained model came to its conclusion
- Training methods involves looking for optimal hyper-parameters rather than contributing human knowledge or intuition.
- No human involvement.
- No contribution to human knowledge.

Interpretable Machine Learning Techniques

- Model gives us information about why new data may have been sorted the way they did.
- Human involvement

Theoretical Machine Learning Research

- Using mathematics to better understand behavior of black box machine learning models.
- Bayesian models - how can we "inject" prior knowledge into training.

k-means Clustering

Intuition

- Aims to partition n observations into k clusters.
- Minimizes within-cluster variances (squared Euclidean distances)

k-means Clustering

Intuition

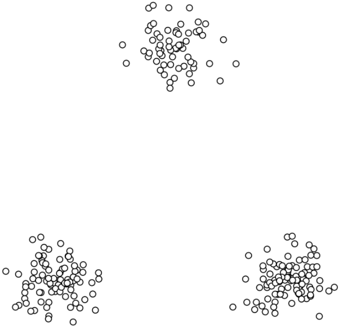
- Aims to partition n observations into k clusters.
- Minimizes within-cluster variances (squared Euclidean distances)

Algorithm

1. Randomly select k center points for each of the k clusters.
2. Assign each data point to the cluster with the closest center.
3. Update cluster centers to minimize squared distances between center and data points.
4. Repeat steps 2-3 until convergence.

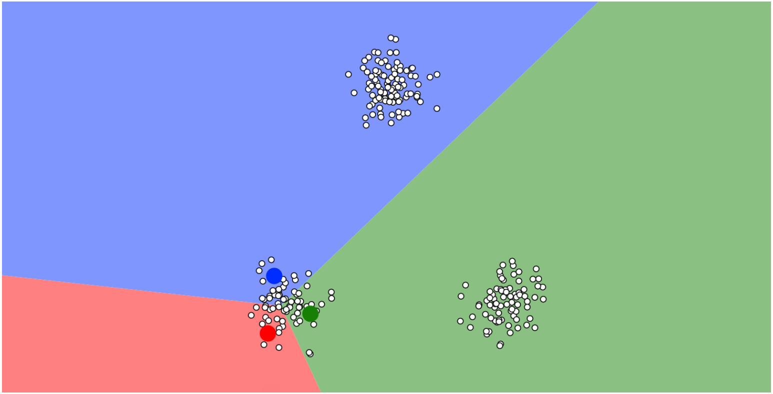
k-means Example

Dataset: Generic 3 clusters

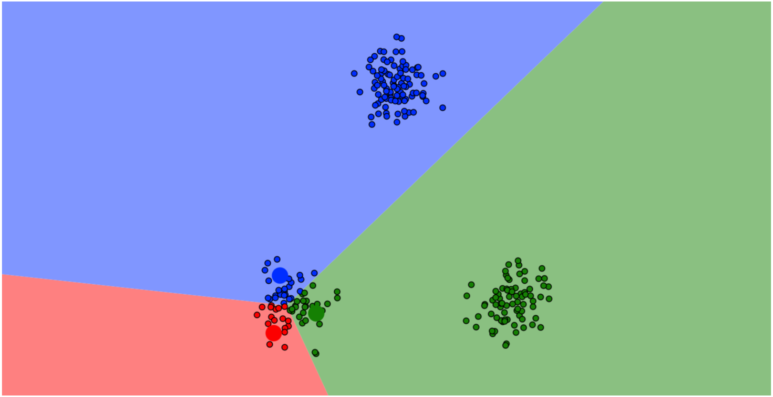


k-means Example

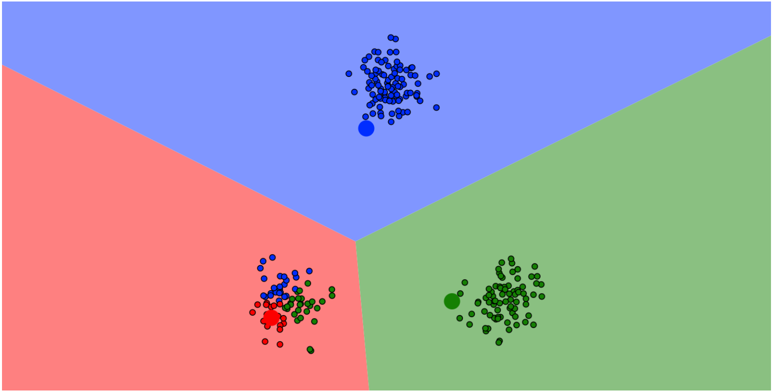
Select $k = 3$ center points for $k = 3$ clusters.



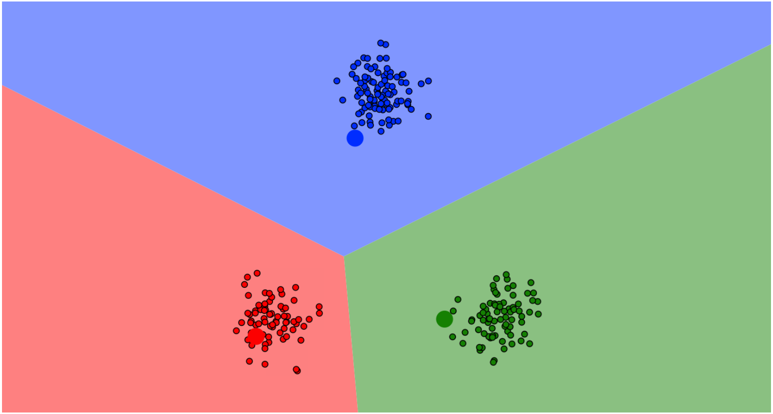
Assign each data point to the cluster with the closest center.



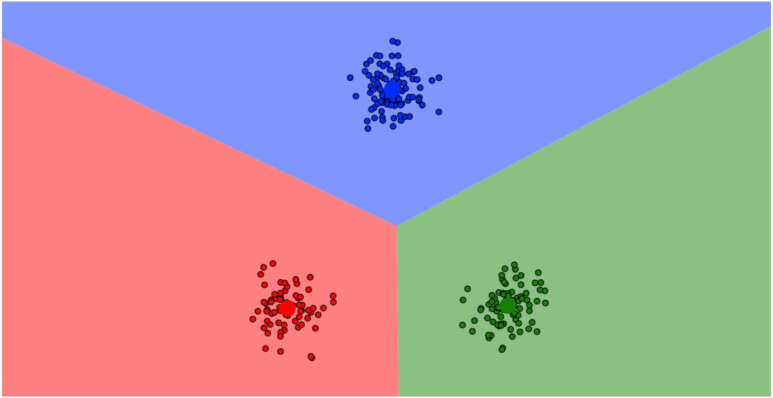
Update cluster centers to minimize squared distances between center and data points.



Assign each data point to the cluster with the closest center.



Update cluster centers to minimize squared distances between center and data points.



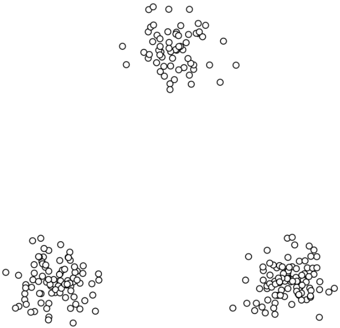
k-means Clustering: Issue 1

Issue 1

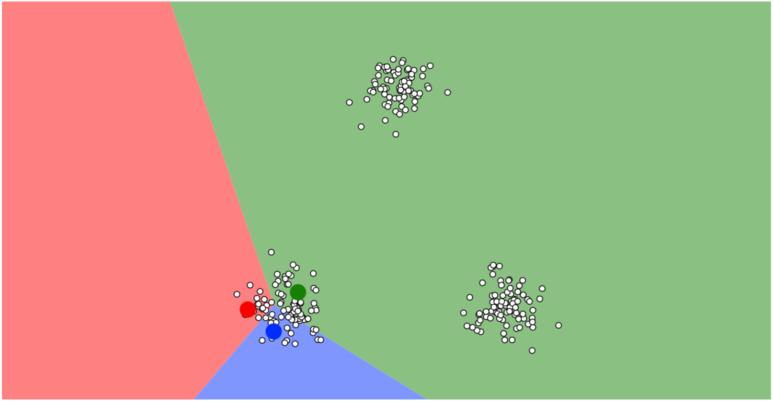
Non-deterministic algorithm

- Initial random assignment of cluster centers
- Also an issue in some models that use stochastic gradient descent for optimization.
  - End up with different models as parameters get stuck in different local minima.

Dataset: Generic 3 clusters

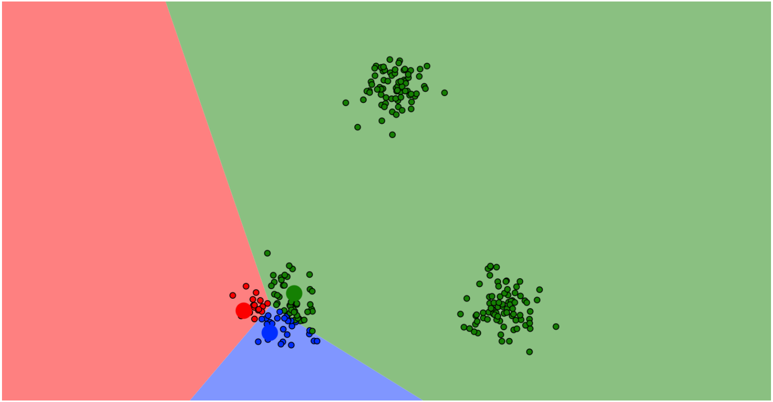


Select $k = 3$ center points for $k = 3$ clusters.



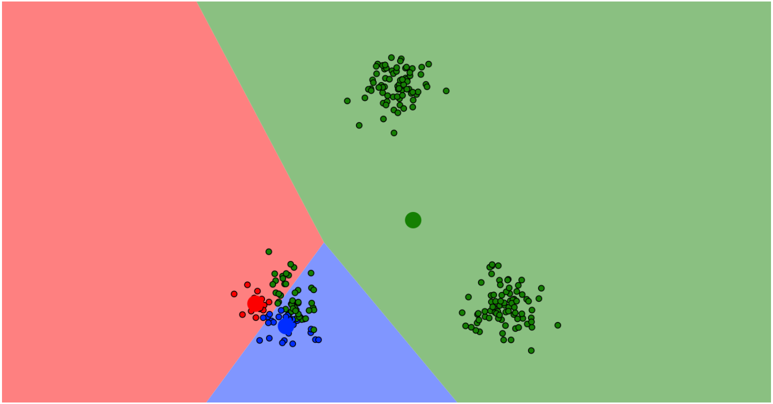
k-means Clustering: Issue 1

Assign each data point to the cluster with the closest center.



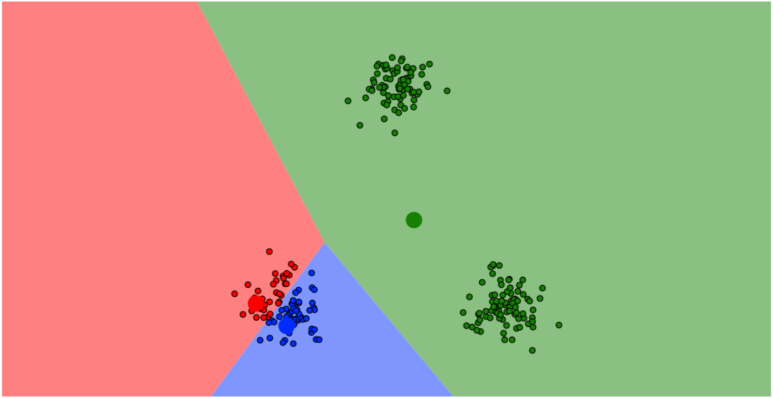
k-means Clustering: Issue 1

Update cluster centers to minimize squared distances between center and data points.

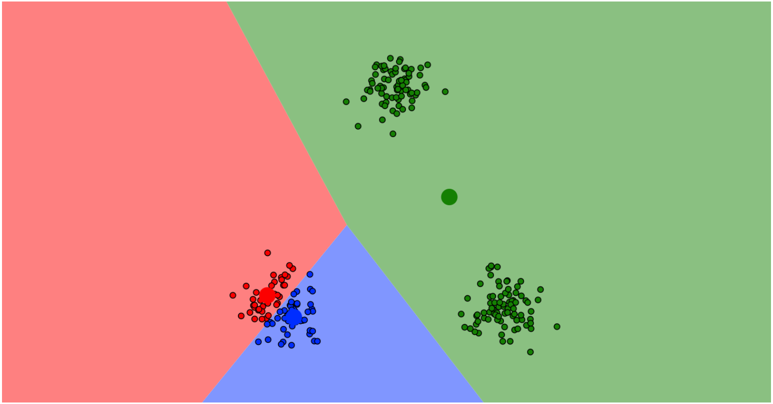


k-means Clustering: Issue 1

Assign each data point to the cluster with the closest center.



Update cluster centers to minimize squared distances between center and data points.

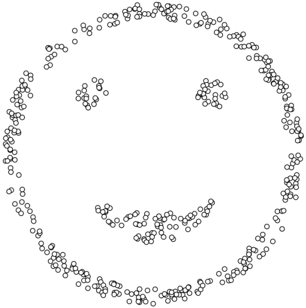


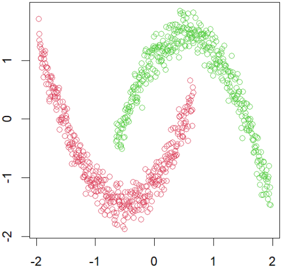
Issue 2

Reliance on Euclidean Distance

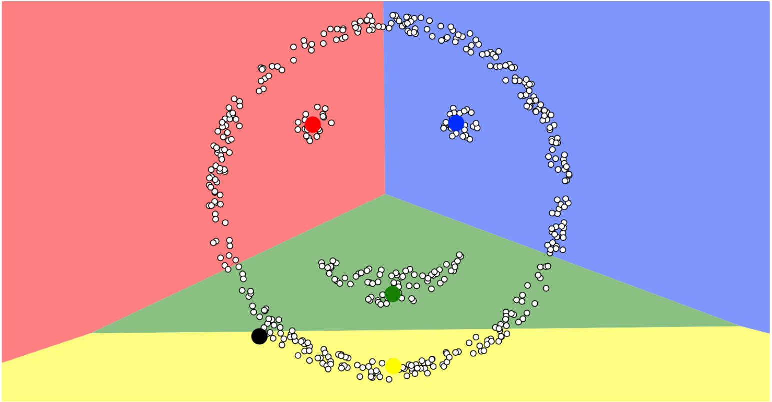
- Problems clustering data sets where groups are not equally distributed around a center
- In high dimensions, notion of distance breaks down.

Datasets: Smiley Face and Half-Moons



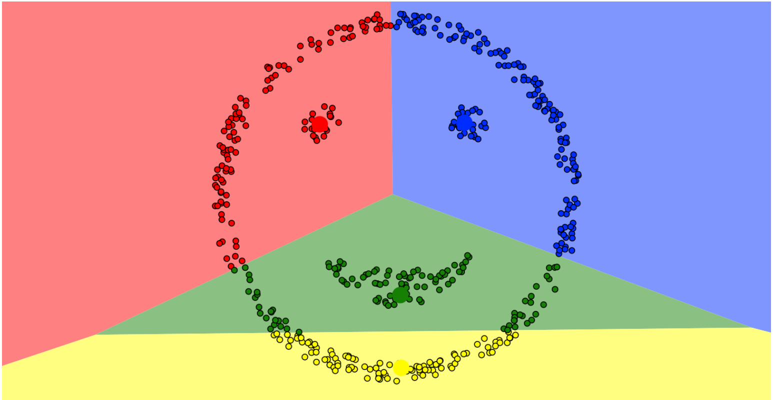


Select $k = 4$ center points for $k = 4$ clusters.



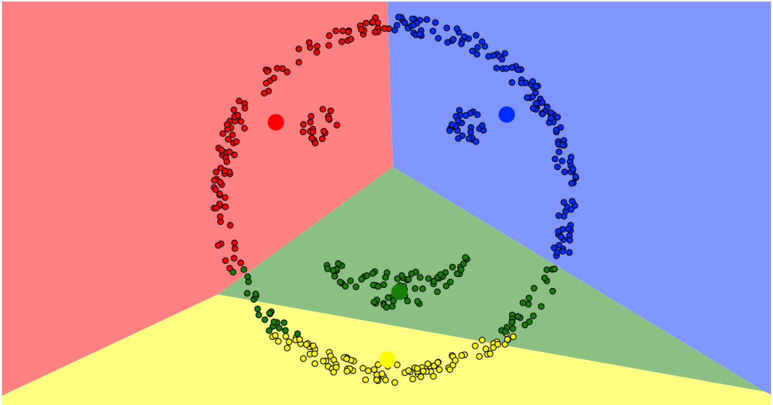
k-means Clustering: Issue 2

Assign each data point to the cluster with the closest center.



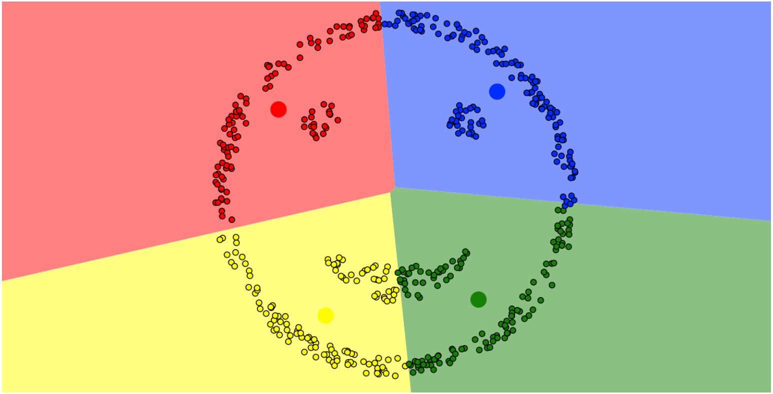
k-means Clustering: Issue 2

Update cluster centers to minimize squared distances between center and data points.



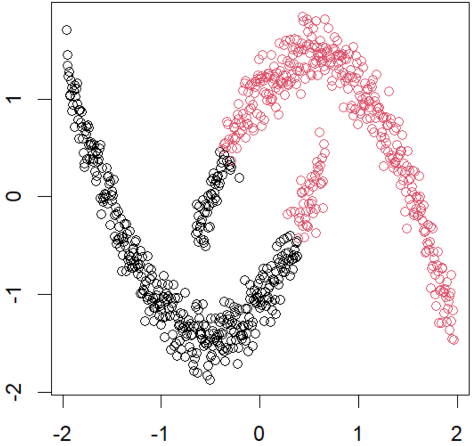
k-means Clustering: Issue 2

Assign each data point to the cluster with the closest center.



k-means Clustering: Issue 2

k-means result of the half-moons dataset.



Statistical Depth

- Methods of ordering multivariate data according to their centrality in a high dimensional data cloud.

Statistical Depth

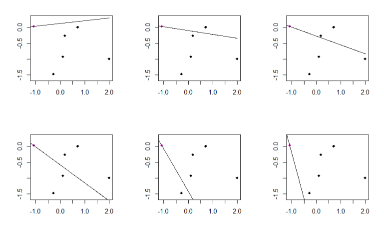
- Methods of ordering multivariate data according to their centrality in a high dimensional data cloud.

Tukey Half-space Depth (Tukey (1974))

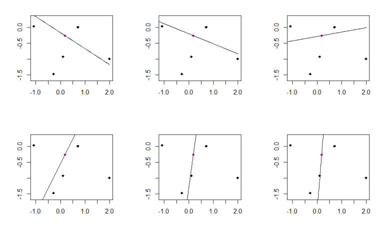
The *Tukey half-space depth* of a point $\mathbf{p} \in \mathbb{R}^d$ with respect to n data points $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n} \in \mathbf{R}^d$ is k/n where k is the smallest number of points in any closed half-space that contains p.

Example Calculation: Tukey Half-space Depth

Dataset: $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_6})$



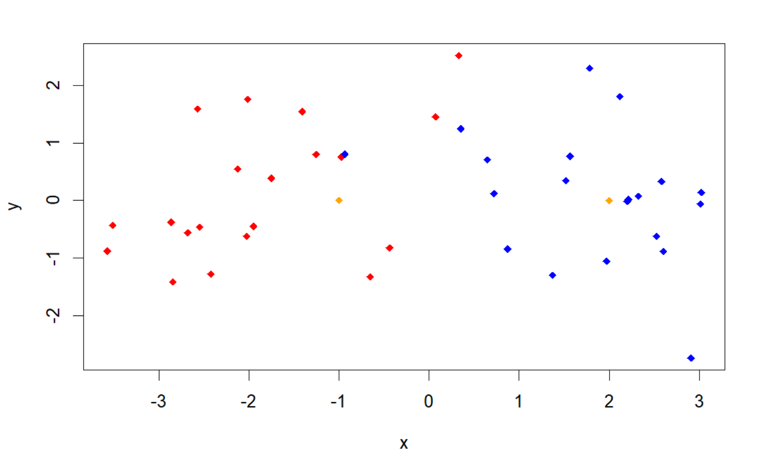
$\mathsf{HD}(x_1|\mathbf{X}) = \frac{1}{6}$



$\mathsf{HD}(x_2|\mathbf{X}) = \frac{2}{6} = \frac{1}{3}$

Classification using Depth

Brief Example in Supervised Learning



Example Calculation: Depth Quantile Functions

Idea/Algorithm for 2 dimensions

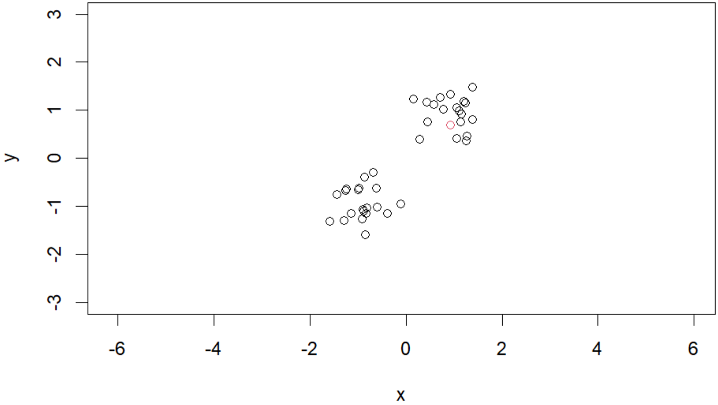
To find the Depth Quantile Function of $x \in \{x_1, x_2, \ldots, x_n\}$

1. Measure "Conal" depths of the midpoint between x and all other points x_i at 100 positions along the line that intersects each pair of points. This results in $(n-1)$ 100 depth values.
2. Sort the 100 depth values. Repeat for all $n-1$ 100-depth vectors called quantile functions.
3. Average $n-1$ quantile vectors into a depth quantile function.
4. Repeat for all n points, which results in n depth quantile functions, one for each point.

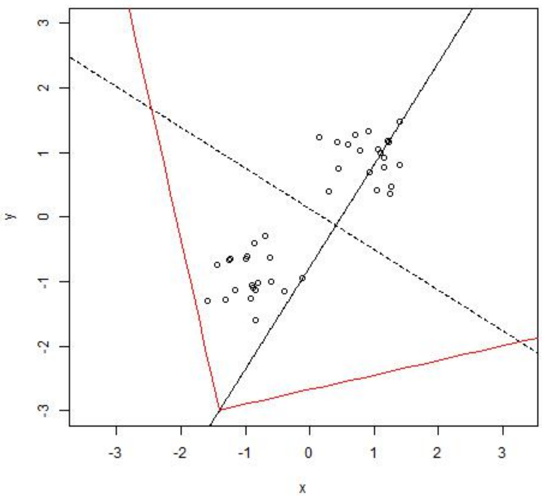
Applies to $nD, n > 2$ space as well.

Example Calculation: Depth Quantile Functions

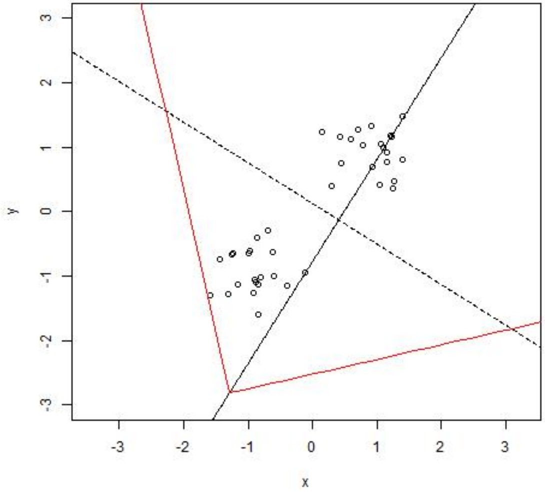
Example Dataset



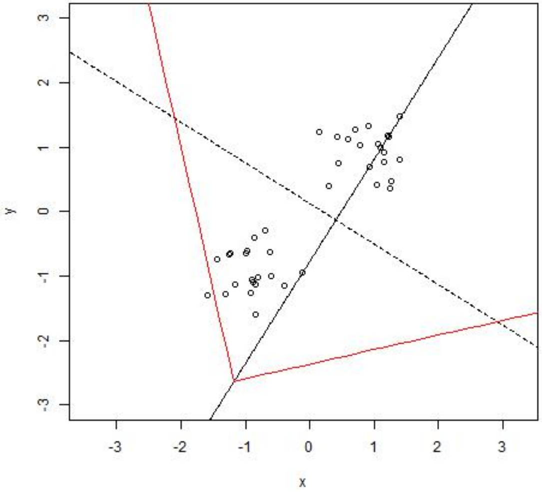
Example Calculation: Depth Quantile Functions



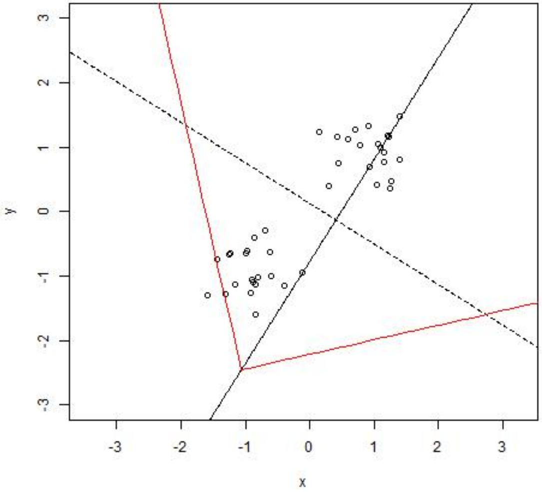
Example Calculation: Depth Quantile Functions



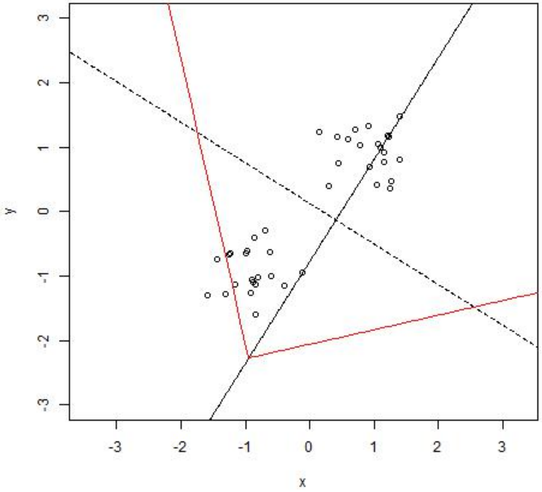
Example Calculation: Depth Quantile Functions



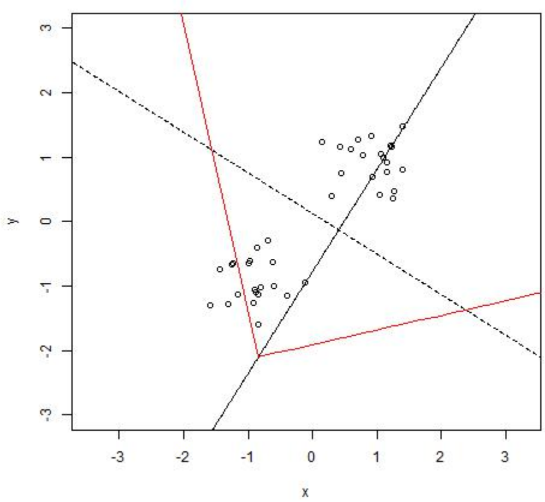
Example Calculation: Depth Quantile Functions



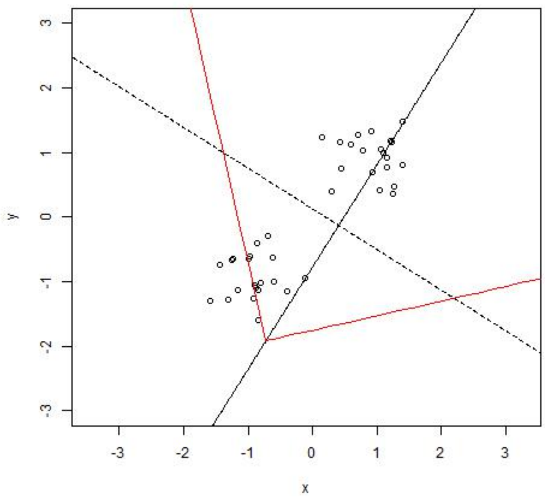
Example Calculation: Depth Quantile Functions



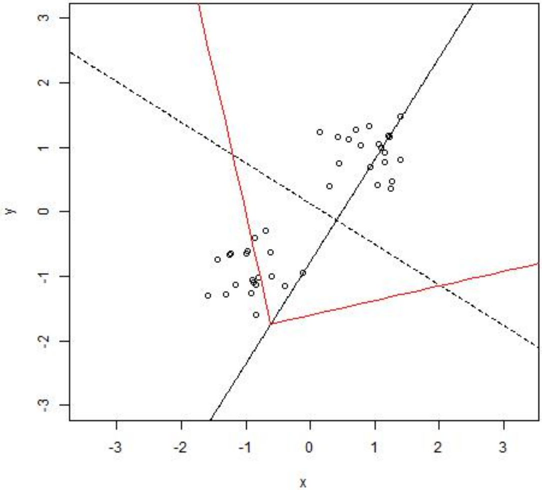
Example Calculation: Depth Quantile Functions



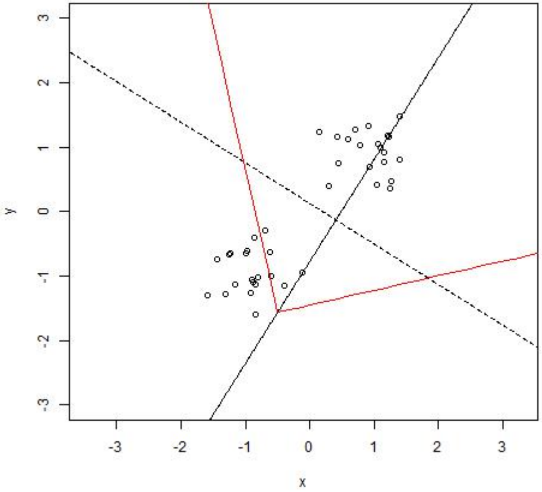
Example Calculation: Depth Quantile Functions



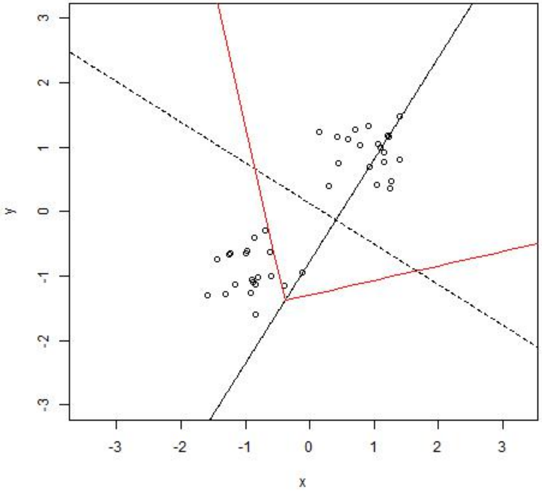
Example Calculation: Depth Quantile Functions



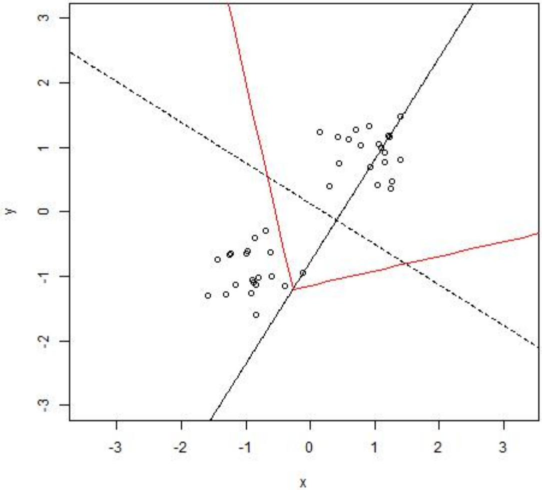
Example Calculation: Depth Quantile Functions



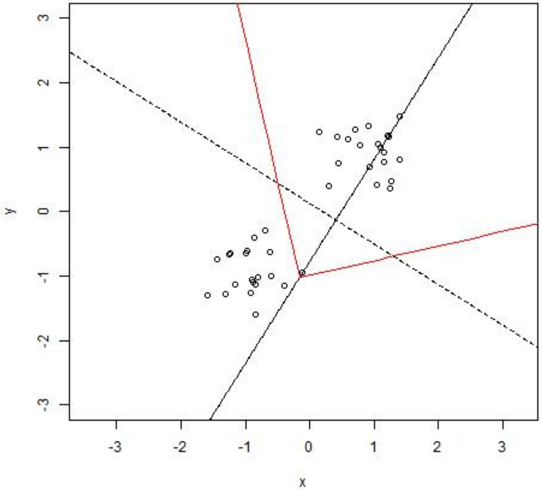
Example Calculation: Depth Quantile Functions



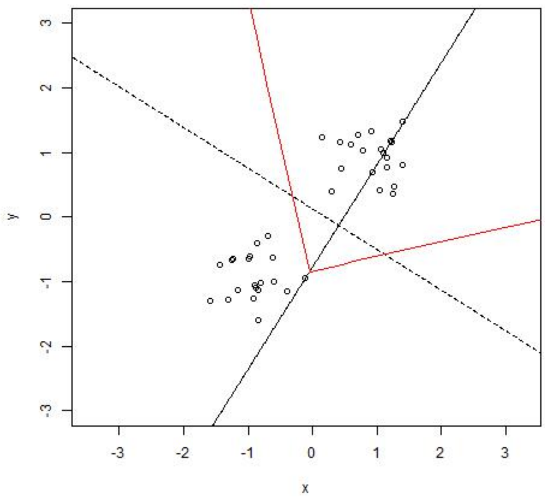
Example Calculation: Depth Quantile Functions



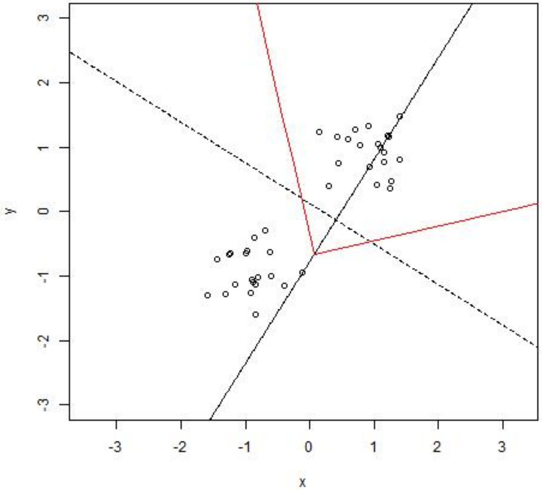
Example Calculation: Depth Quantile Functions



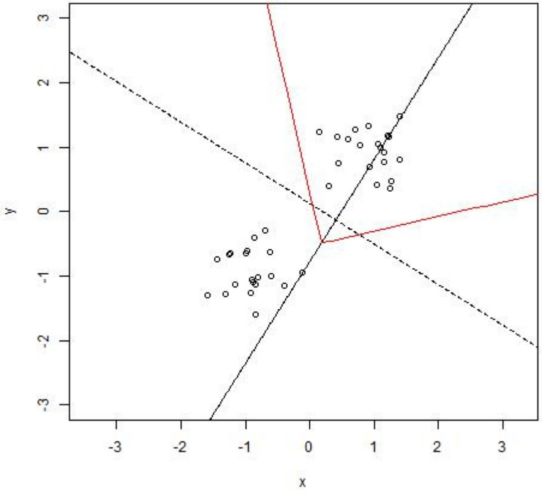
Example Calculation: Depth Quantile Functions



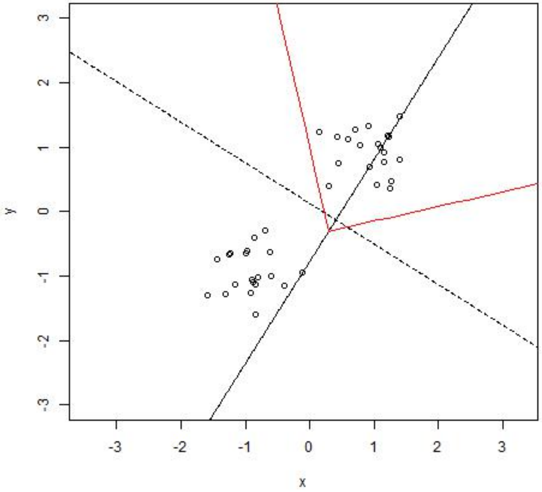
Example Calculation: Depth Quantile Functions



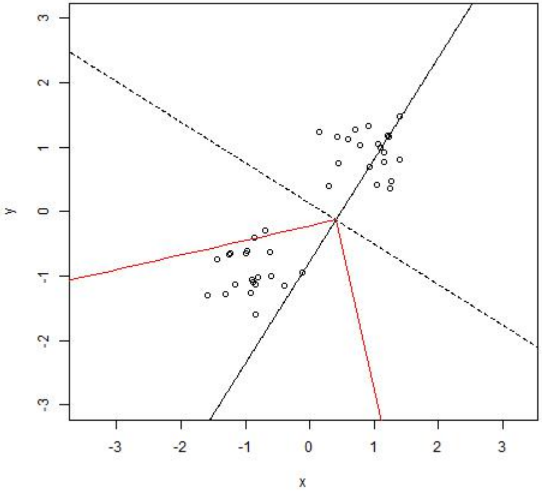
Example Calculation: Depth Quantile Functions



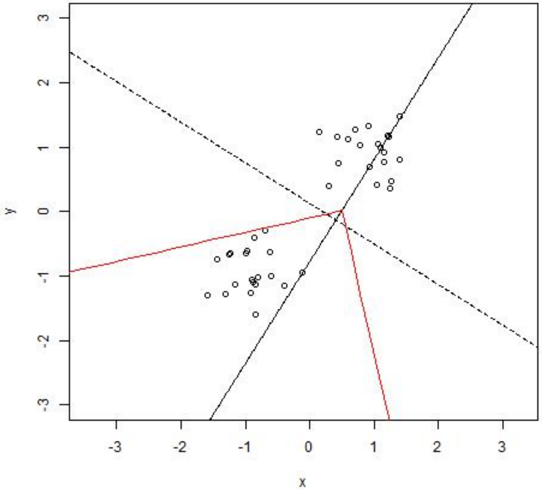
Example Calculation: Depth Quantile Functions



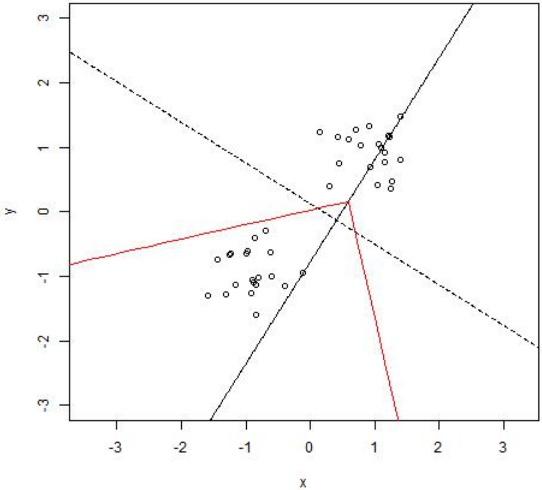
Example Calculation: Depth Quantile Functions



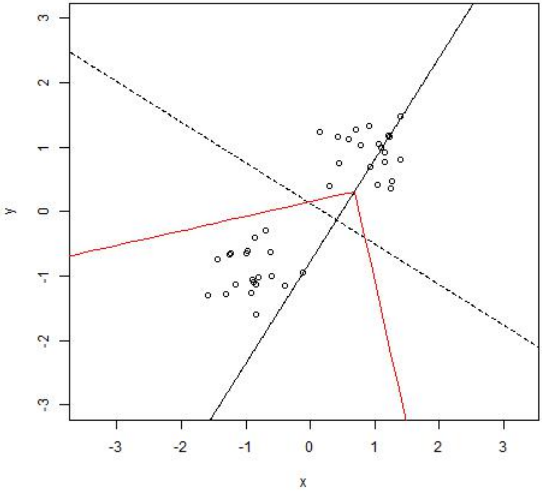
Example Calculation: Depth Quantile Functions

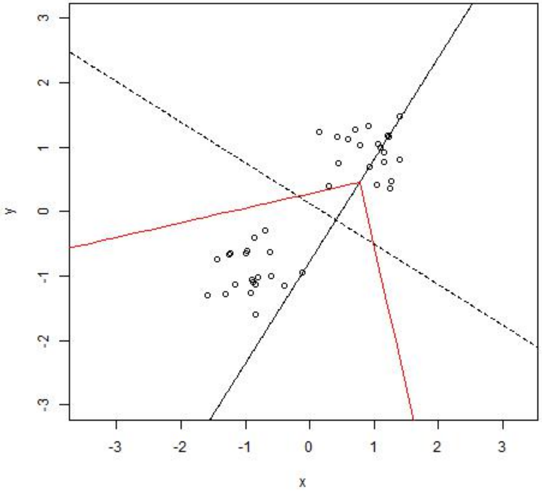


Example Calculation: Depth Quantile Functions



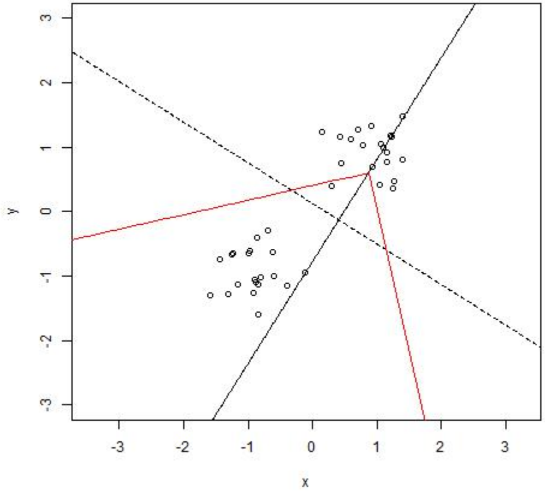
Example Calculation: Depth Quantile Functions



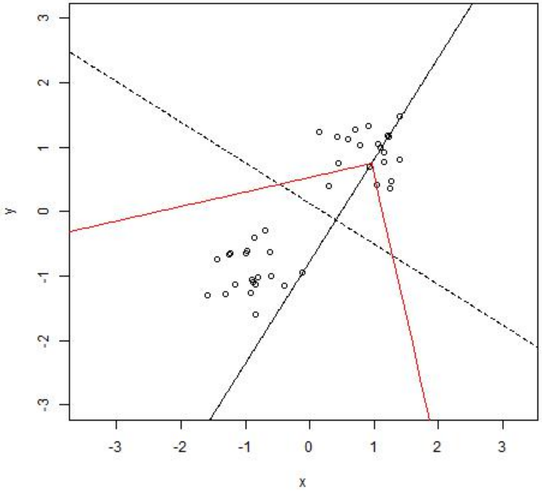


3
2
1
0
-1
-2
-3
y
-3
-2
-1
0
1
2
3
x

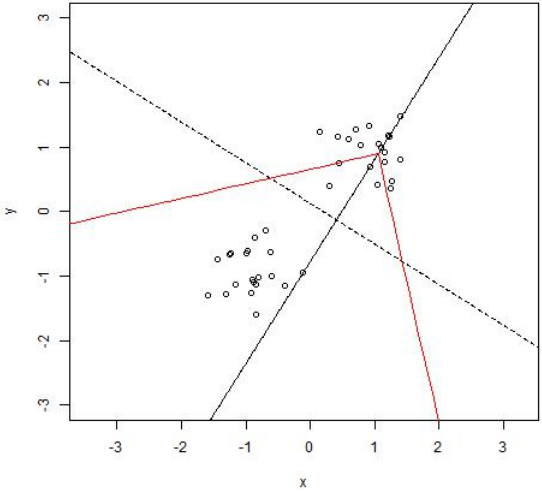
Example Calculation: Depth Quantile Functions



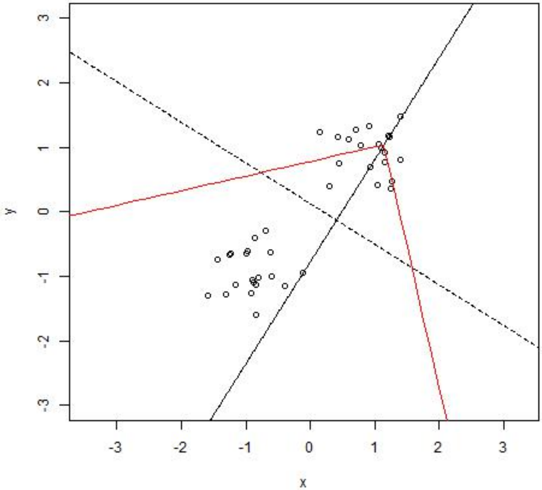
Example Calculation: Depth Quantile Functions



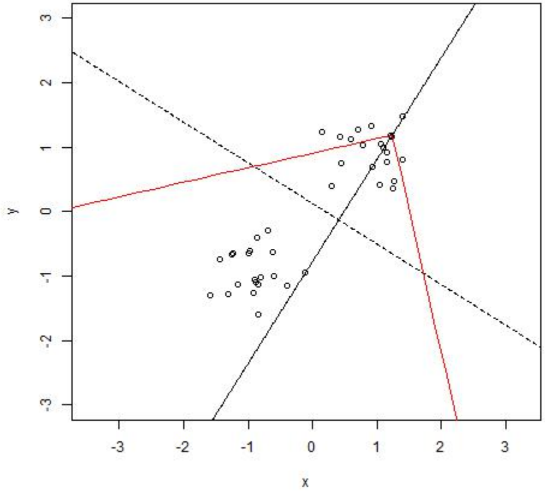
Example Calculation: Depth Quantile Functions



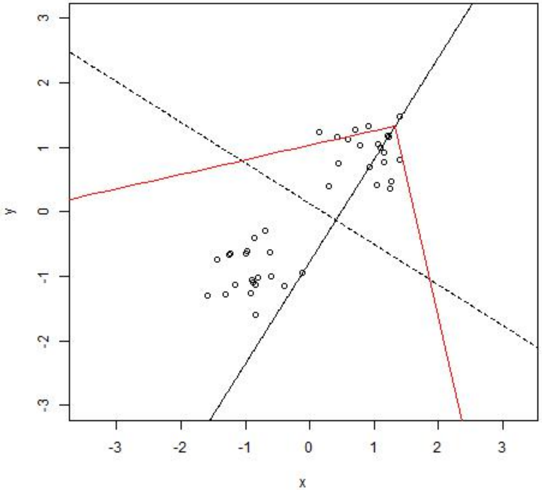
Example Calculation: Depth Quantile Functions



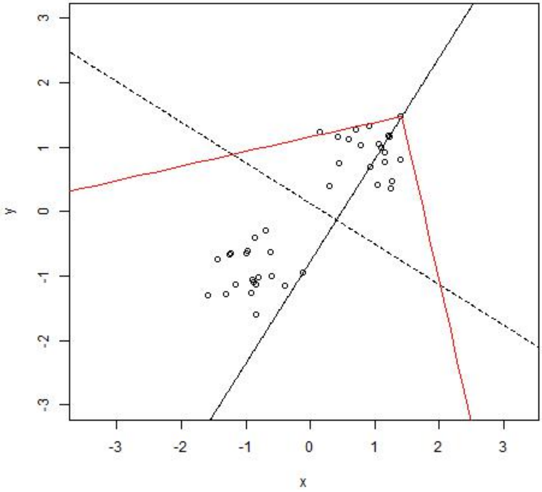
Example Calculation: Depth Quantile Functions



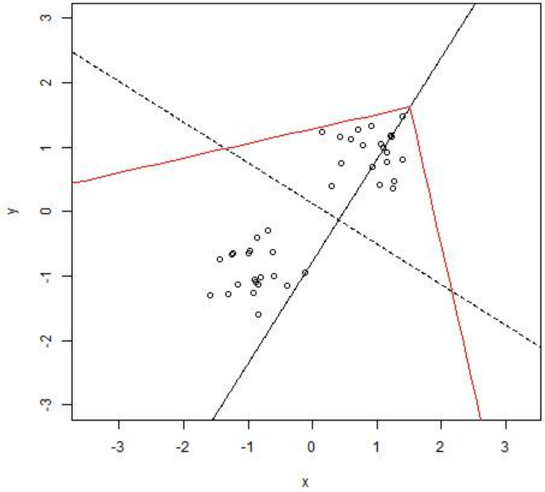
Example Calculation: Depth Quantile Functions



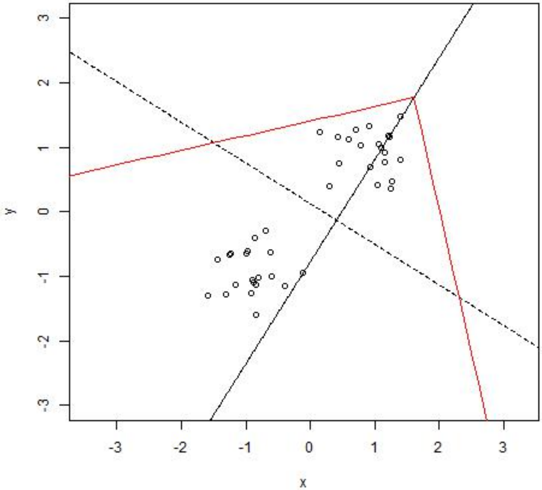
Example Calculation: Depth Quantile Functions



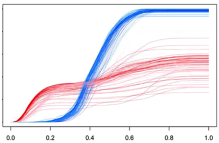
Example Calculation: Depth Quantile Functions



Example Calculation: Depth Quantile Functions



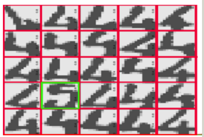
Classification (Chandler, Polonik 2021)

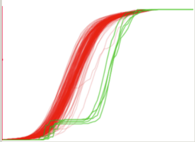


(a) Iris Data: Setosa vs. Versicolor

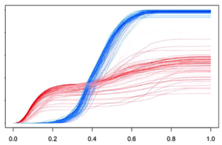
Anomaly Detection (2022)

Multiple Features Dataset
$d = 649$ features of 200 '4's and 5 '5's.





Classification (Chandler, Polonik 2021)

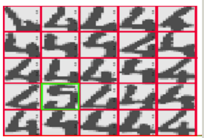


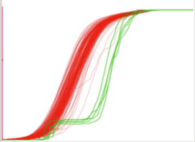
(a) Iris Data: Setosa vs. Versicolor

Anomaly Detection (2022)

Multiple Features Dataset

$d = 649$ features of 200 '4's and 5 '5's.





Properties of Depth Quantile Functions

- (+) Deterministic (for the most part...)
- (+) Does not rely on Euclidean Distance
- (+) Provides visualizations for human understanding and intervention
- (-) computationally intensive

DQF for Clustering

Idea: Perform partial clustering followed by Anomaly Detection.

General Algorithm

1. Partially perform traditional clustering methods such as k-means or hierarchical clustering on data set.
2. Calculate Depth Quantile Function for clusters that are close.
3. Decide whether we should merge the two clusters. (Potential Supervision)

DQF for Clustering

Example: 20-dimension dataset with 802 observations.

X1 <dbl>	X2 <dbl>	X3 <dbl>	X4 <dbl>	X5 <dbl>
-0.358756705	-4.562908248	1.34841914	4.691613336	-1.515818781
-0.392823068	-4.436602127	1.55082440	4.279485056	-1.525100127
-0.407749470	-4.370597977	1.64092653	4.082567137	-1.526601088
-0.421181142	-4.309303095	1.72225836	3.902456026	-1.527494441
-0.412952068	-4.316252113	1.67649644	3.965899048	-1.519582713
-0.444704237	-4.197237045	1.86532324	3.579791646	-1.527923340
-0.443766727	-4.181213414	1.86234434	3.561248093	-1.522975628
-0.420092299	-4.236824250	1.72595728	3.798359196	-1.508785370
-0.428393114	-4.191694538	1.77718397	3.675939601	-1.507593032
-0.460562277	-4.071365700	1.96845746	3.285143511	-1.516103167

1-10 of 802 rows | 1-5 of 20 columns   Previous 1 2 3 4 5 6 ... 81 Next

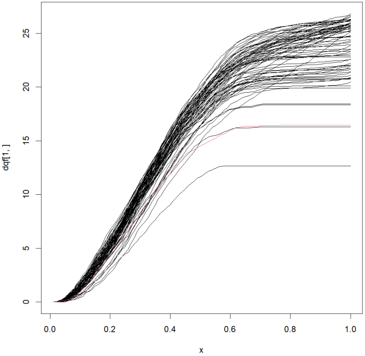
Partially cluster the data set ($k = 7$)

```
```{r}
m.df.hd <- scale(m.df.hd)
dist.m <- dist(m.df.hd, method = 'euclidean')
hc <- hclust(dist.m, method = "average")
fit <- cutree(hc, k = 7)
table(fit)
```
```

```
fit
  1   2   3   4   5   6   7
 73  55 273  85 164 120  32
```

DQF for Clustering

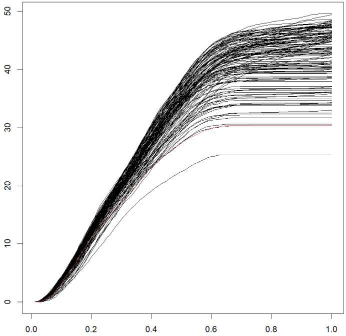
2. Calculate Depth Quantile Functions for the closest point a nearby cluster relative to the cluster
3. Decide whether we should merge the two clusters. (Supervision)



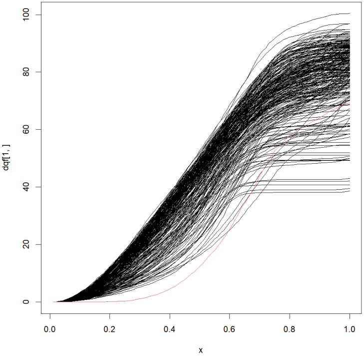
Cluster 1 and Cluster 2.

DQF for Clustering

2. Calculate Depth Quantile Function for the closest point in a nearby cluster.
3. Decide whether we should merge the two clusters. (Supervision)



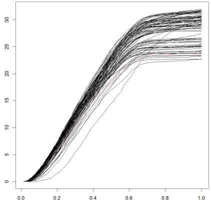
Cluster [1 and 2] and Cluster 3
Merge.



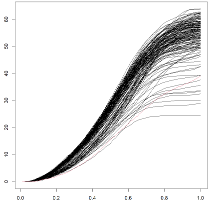
Cluster 3 and Cluster 4.
Do not Merge.

DQF for Clustering

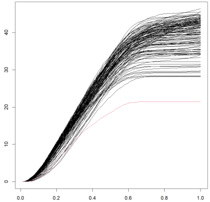
2. Calculate Depth Quantile Function for the closest point in a nearby cluster.
3. Decide whether we should merge the two clusters. (Supervision)



Cluster 4 and Cluster 5
Merge.



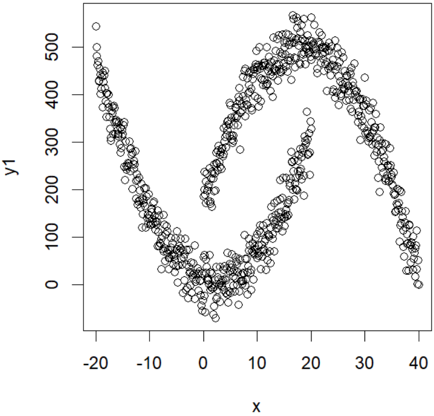
Cluster 5 and Cluster 6.
Merge.



Cluster 6 and Cluster 7.
Merge.

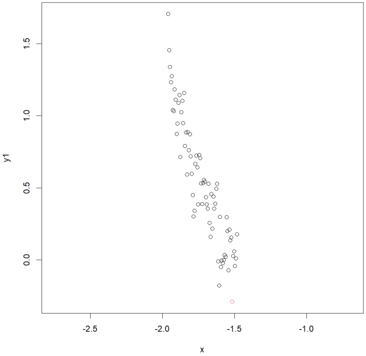
DQF for Clustering

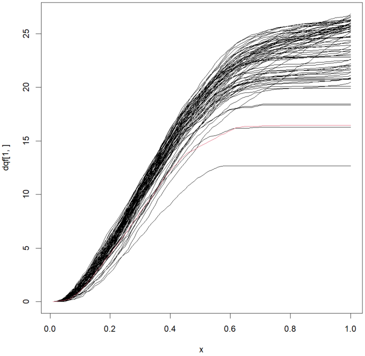
Turns out... it's just the half moon data set raised into 20 dimensions.



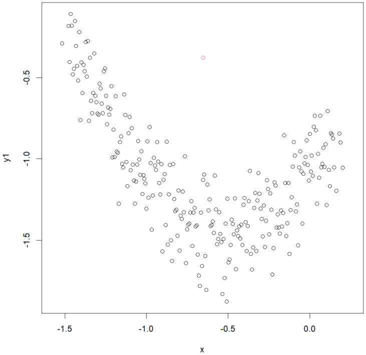
DQF for Clustering

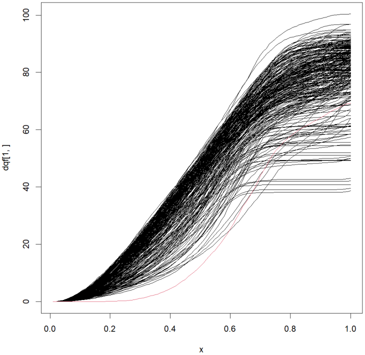
When we decided merged cluster 1 and cluster 2.





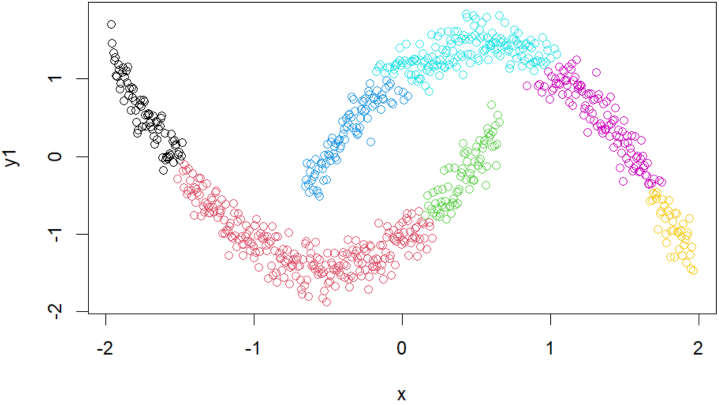
When we decided not to merged cluster 3 and cluster 4.





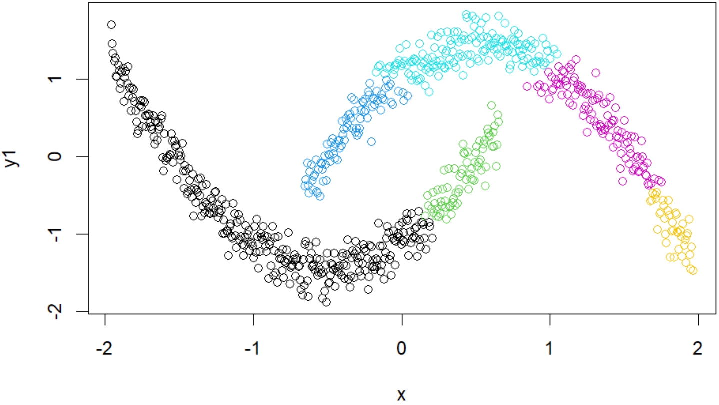
DQF for Clustering

Start with 7 Clusters



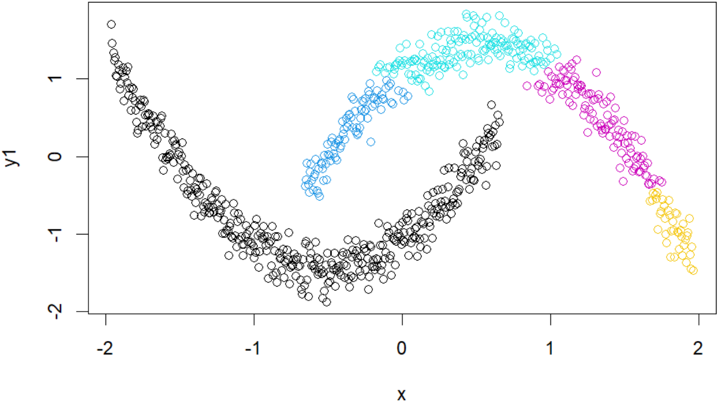
DQF for Clustering

Merge 1 and 2.



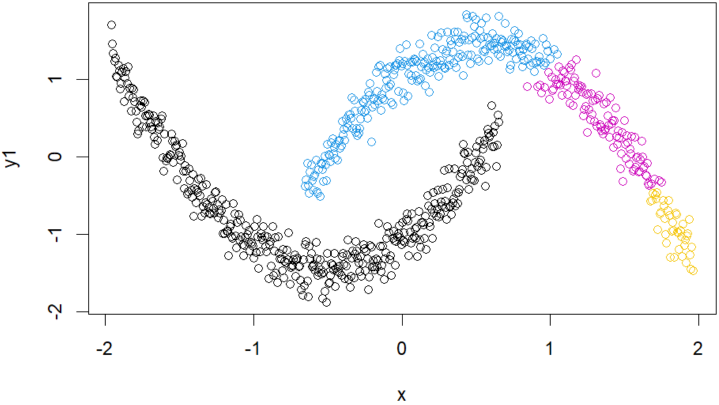
DQF for Clustering

Merge [1 and 2] and 3.



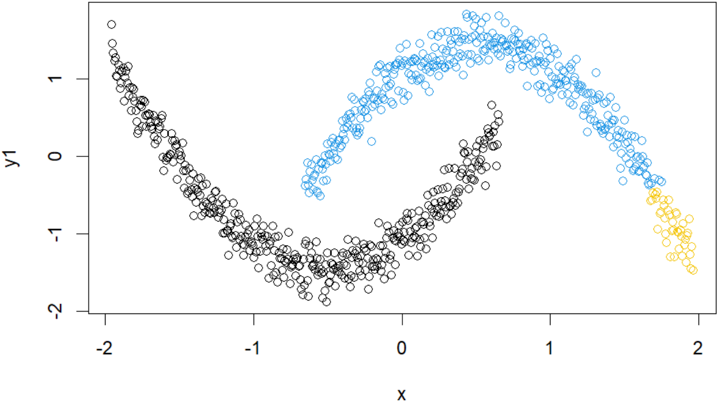
DQF for Clustering

Do not merge 4. Merge 4 and 5.



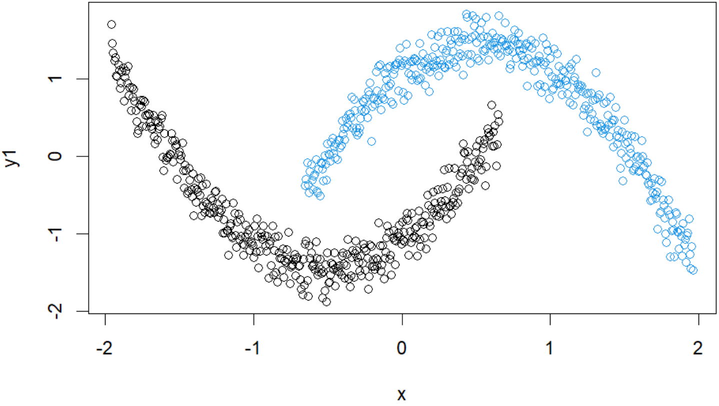
DQF for Clustering

Merge [4 and 5] and 6



DQF for Clustering

Merge [4,5, and 6] and 7.



Future Directions

Now working on...

- Looking for appropriate metrics to:
  - Automatically categorize DQFs to decide whether to merge clusters or not based on similarity of functions.
  - Measure confidence for merging clusters to indicate when human supervision/intervention may be necessary.

Future Directions

Now working on...

- Looking for appropriate metrics to:
    - Automatically categorize DQFs to decide whether to merge clusters or not based on similarity of functions.
    - Measure confidence for merging clusters to indicate when human supervision/intervention may be necessary.
- Writing.

Thank you to Professor Chandler!

Thank you to my peers and professors in the Math department and to my friends.