Universidad de Guadalajara

Thompson Rivers University Seminar

Introduction to Machine Learning

Dr. Carlos Villaseñor

Content

Day 2 (Dr. Carlos Villaseñor)

- Classification problem
- Logistic regression
- Classification metrics
  - Confusion matrix
  - Classification report

Rest

- Nonlinear classification (DT, SVM, KNN, MLP)
- Practice Classification

Rest

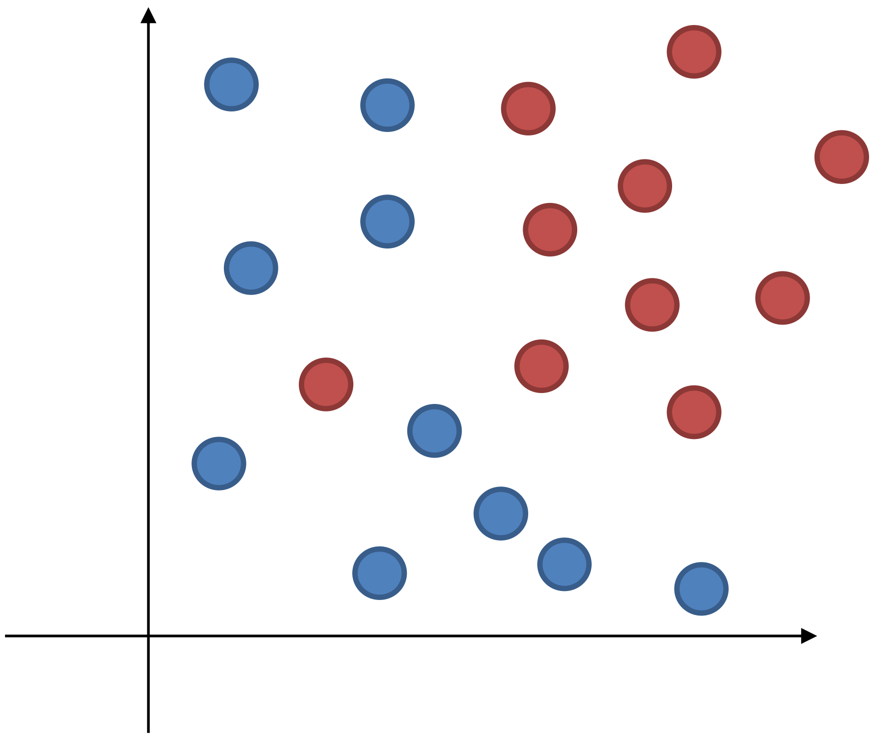
- Non supervised learning
  - Clustering techniques (K-means, Spectral clustering, DBSCAN)
  - Silhouette score
  - Dimensionality reduction (PCA, t-SNE)

Course repository

https://github.com/Dr-Carlos-Villasenor/TRSeminar.git

Classification problem

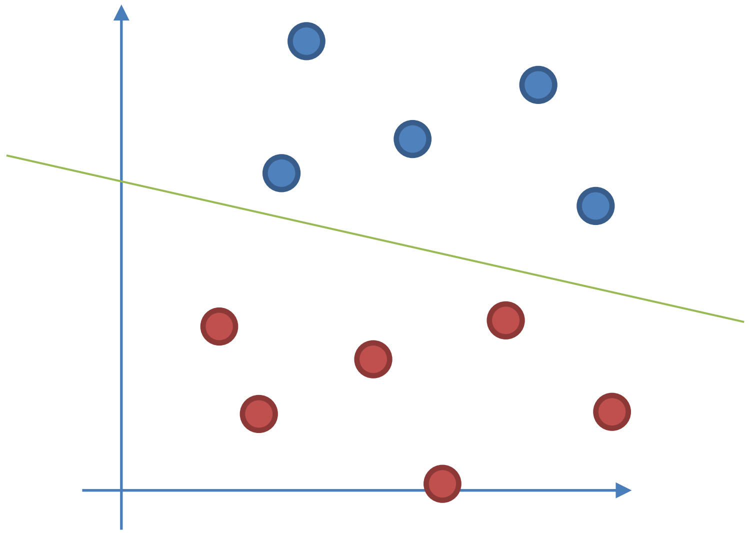
The classification problem is the supervised problem with a categorical output



Logistic regression

- What if instead of a class, we wanted to have a probability

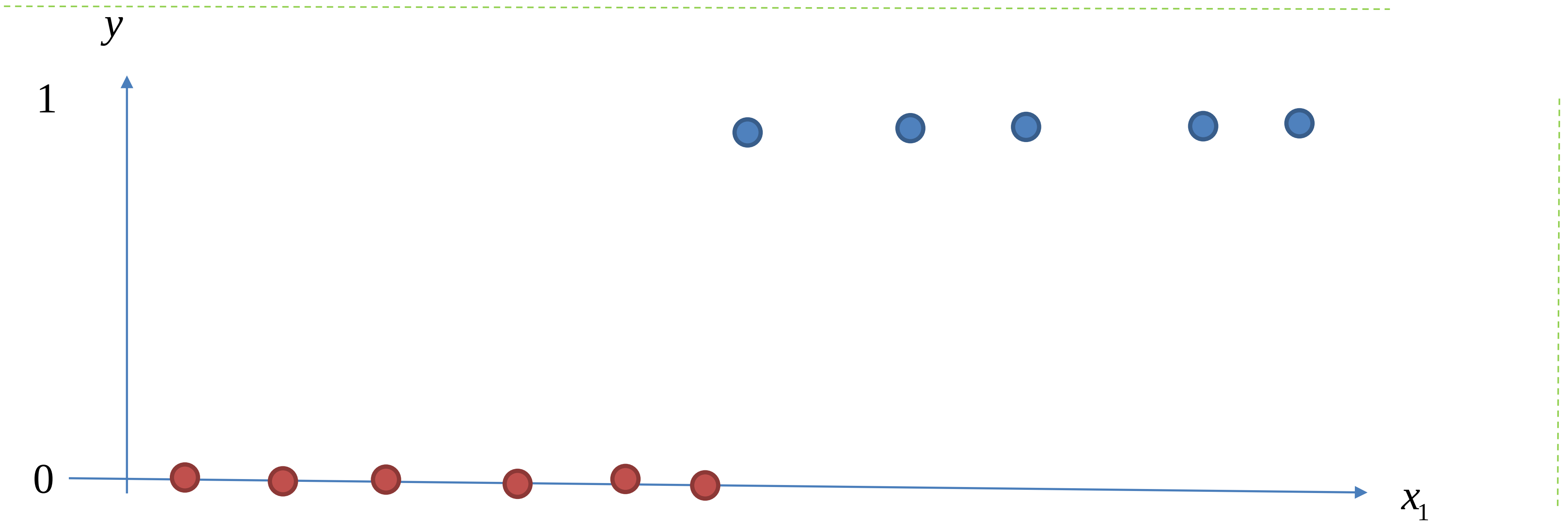
$$w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$



Logistic regression

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$$

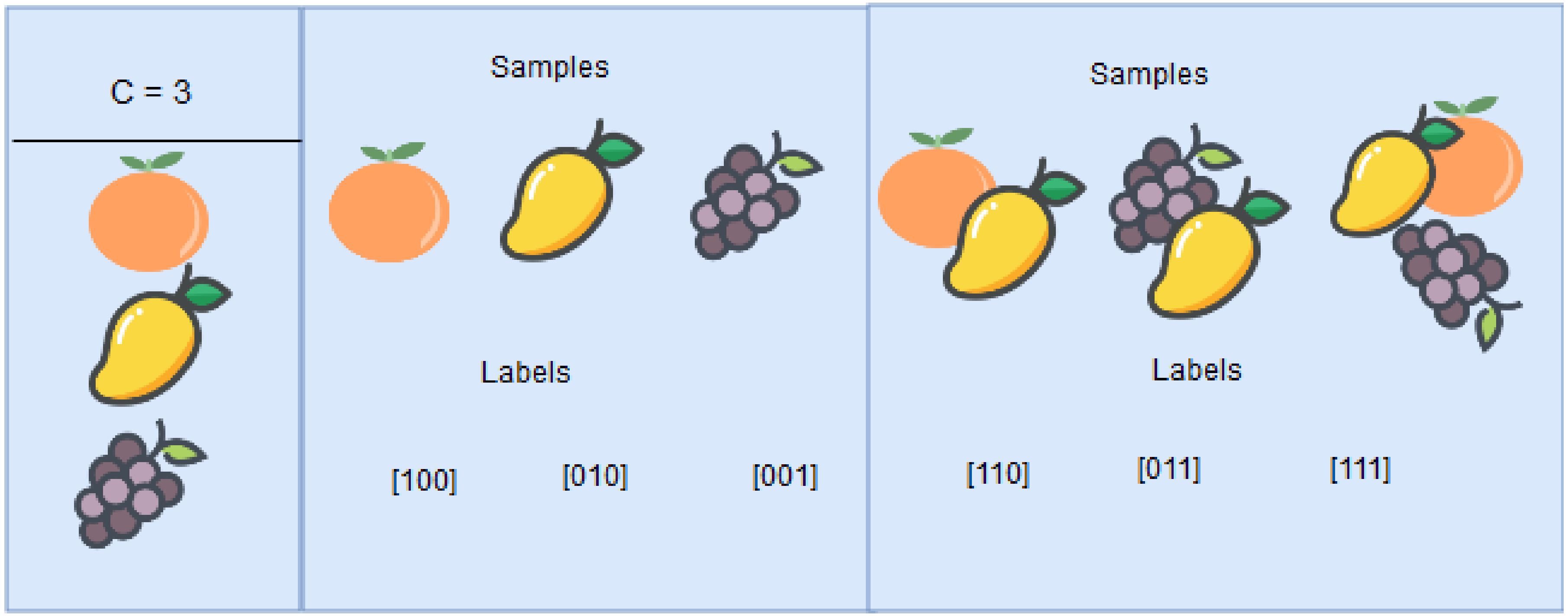
$$p = \frac{1}{1+e^{-z}}$$



Let's get hands on! Open https://github.com/Dr-Carlos-Villasenor/TRSeminar/blob/main/TRS07_LogisticRegression.ipynb

Types of Classification

Multi-Class **Multi-Label**

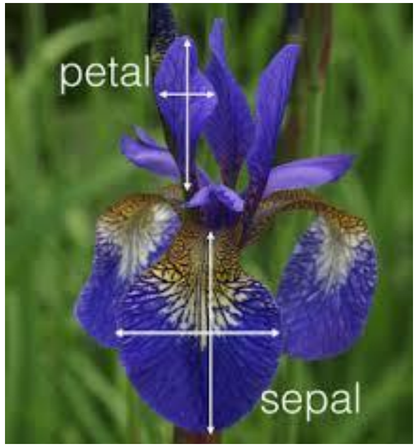


$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} 0.88 \\ 0.12 \\ 0.23 \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} 0.79 \\ 0.11 \\ 0.74 \end{bmatrix}$$

Multi-class classification

- This flower has 3 species
- Iris Setosa
- Iris Versicolour
- Iris Virginica
  - 1. Sepal length in cm
  - 2. Sepal width in cm
  - 3. Petal length in cm
  - 4. Petal width in cm



One vs All

Iris Setosa
Iris Versicolour
Iris Virginica

$\longrightarrow$

1
2
3

$$\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$v_i = \begin{cases}1 & i = c\\ 0 & i \neq c\end{cases}$$

Binary Confusion Matrix

		Predictions	
		0	1
True Values	0	True Negatives TN	False Negatives FP Type I Error
	1	False Negative FN Type II Error	True Positives TP

Example

y	0	1	0	1	0	0	1	0	0	1	0	1	1	0	1
$\hat{y}$	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1

		Predictions	
		0	1
True Values	0		
	1		

Some metrics

- Accuracy is an overall performance metric for the classifier

$$\text{Acc}=\frac{TP+TN}{TP+TN+FP+FN}$$

- In our example

$$\text{Acc}=\frac{5+7}{5+7+1+2}=0.8$$

Some metrics per class

- Class 1 Precision

$$P_1 = \frac{TP}{TP + FP}$$

- In our example

$$P_1 = \frac{5}{5+1} = 0.8333$$

Some metrics per class

- Class 0 Precision

$$P_0 = \frac{TN}{TN + FN}$$

- In our example

$$P_0 = \frac{7}{7+2} = 0.7777$$

Some metrics per class

- Class 1 Recall

$$R_1 = \frac{TP}{TP + FN}$$

- In our example

$$R_1 = \frac{5}{5+2} = 0.7142$$

Some metrics per class

- Class 0 Recall

$$R_0 = \frac{TN}{TN + FP}$$

- In our example

$$R_1 = \frac{7}{7+1} = 0.875$$

Some metrics per class

- F1-score

$$F1_i = 2\frac{P_i R_i}{P_i + R_i}$$

- For Class 1 & 0

$$F1_1 = 2\frac{(0.8333)(0.7142)}{(0.8333)+(0.7142)} = 0.7693$$

$$F1_0 = 2\frac{(0.7777)(0.875)}{(0.7777)+(0.875)} = 0.8238$$

F1-score macro average

- F1-score macro average is an overall measure of a classifier's performance, it is the weighted combination of each

$$\text{F1-macro} = \sum_{i=1}^{c} \frac{|c|}{|D|} \text{F1}_i$$

- In our example:

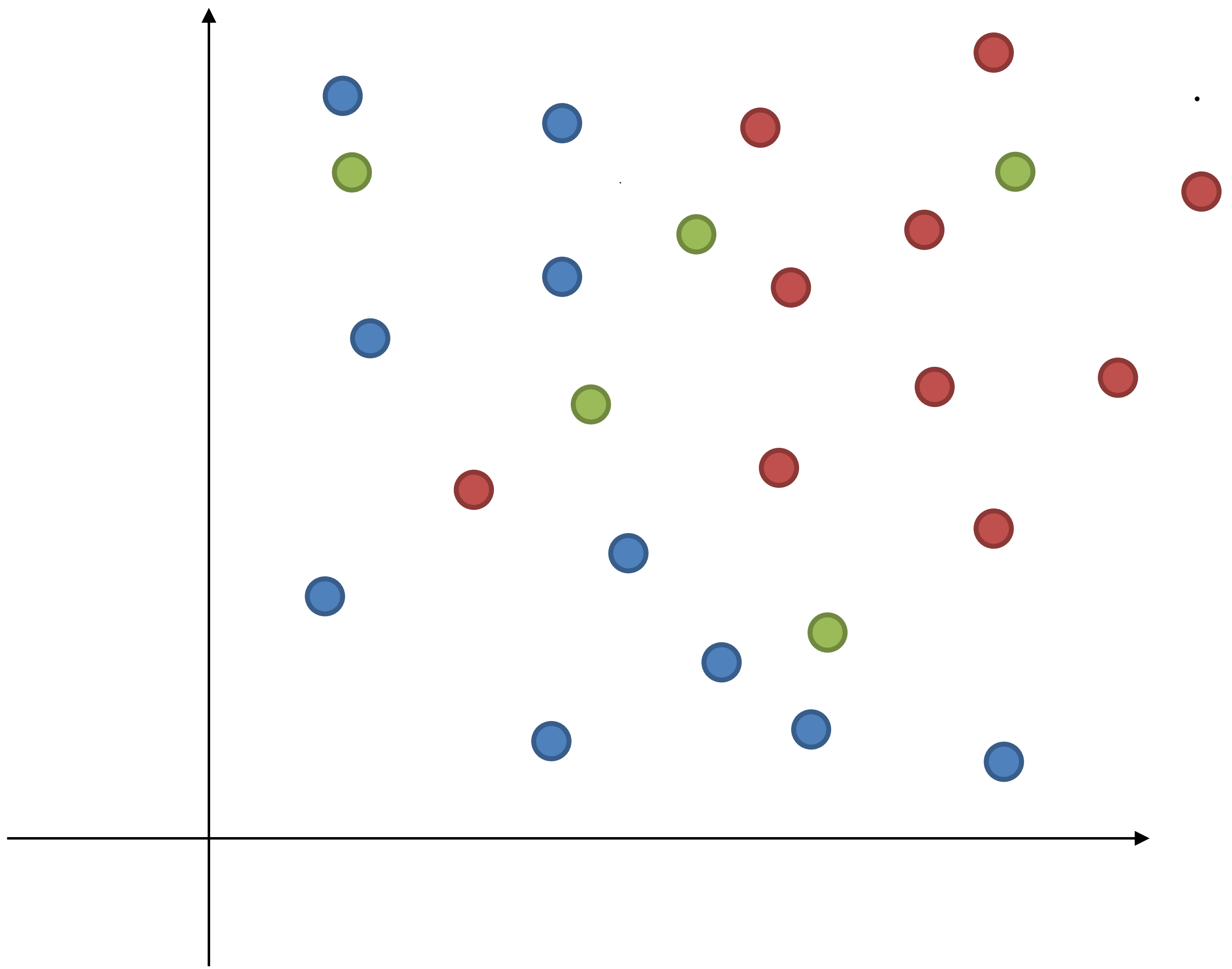
$$\text{F1-macro} = \frac{8}{15}\text{F1}_0 + \frac{7}{15}\text{F1}_1 = \frac{8}{15}0.8238 + \frac{7}{15}0.7693 = 0.7983$$

Torta ahogada

Tota ahogada (*drowned submarine sandwich*) is a typical dish from Jalisco.

Let’s take a rest!

K-NN for classification



Let's get hands on! Open https://github.com/Dr-Carlos-Villasenor/TRSeminar/blob/main/TRS08_Non_linear_classifiers.ipynb

Different metrics

Manhattan Distance (L1)

$$d(p,q) = \sum_i |p_i - q_i|$$

Euclidean distance (L2)

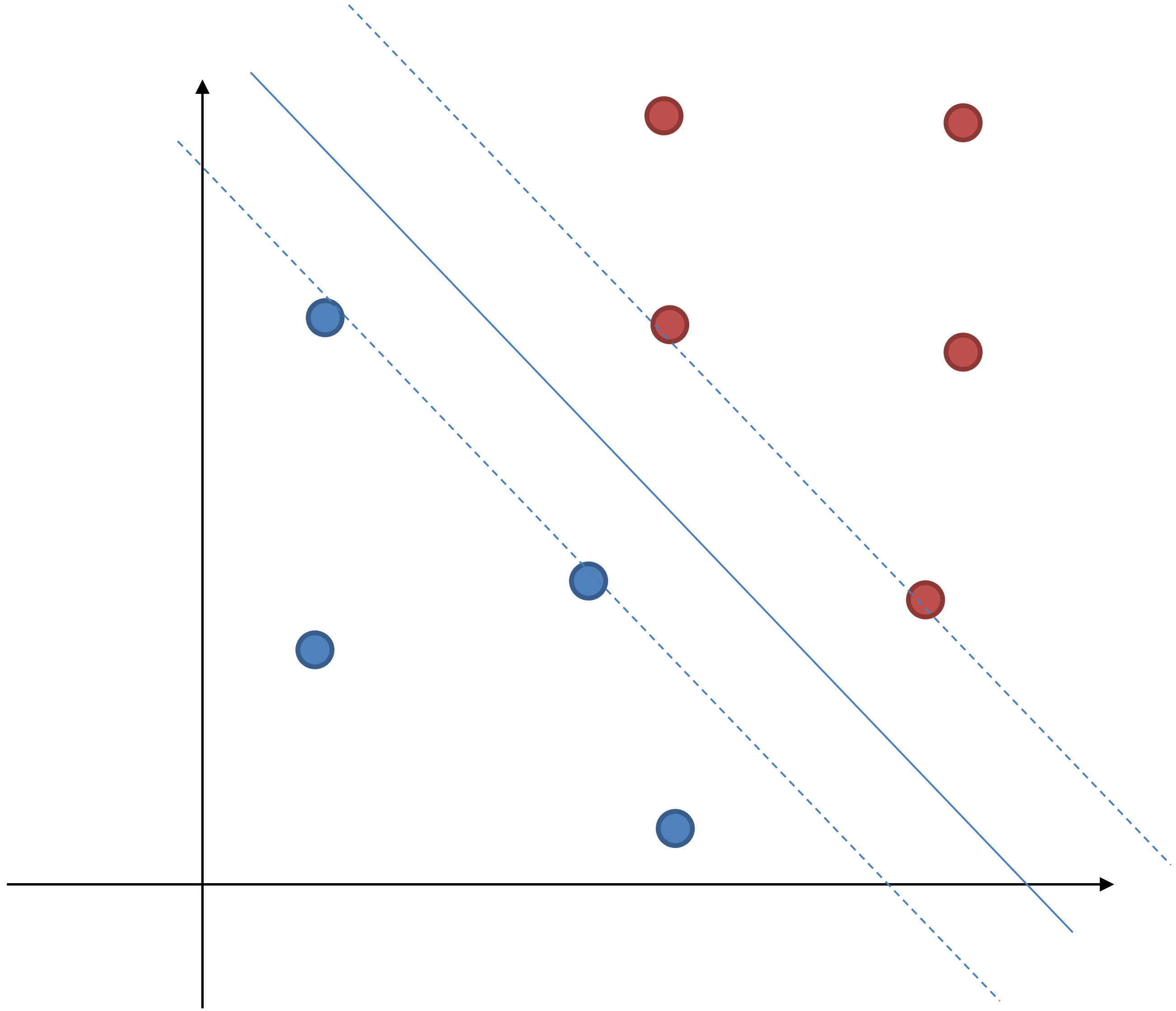
$$d(p,q) = \sqrt{\sum_i (p_i - q_i)^2}$$

Different metrics

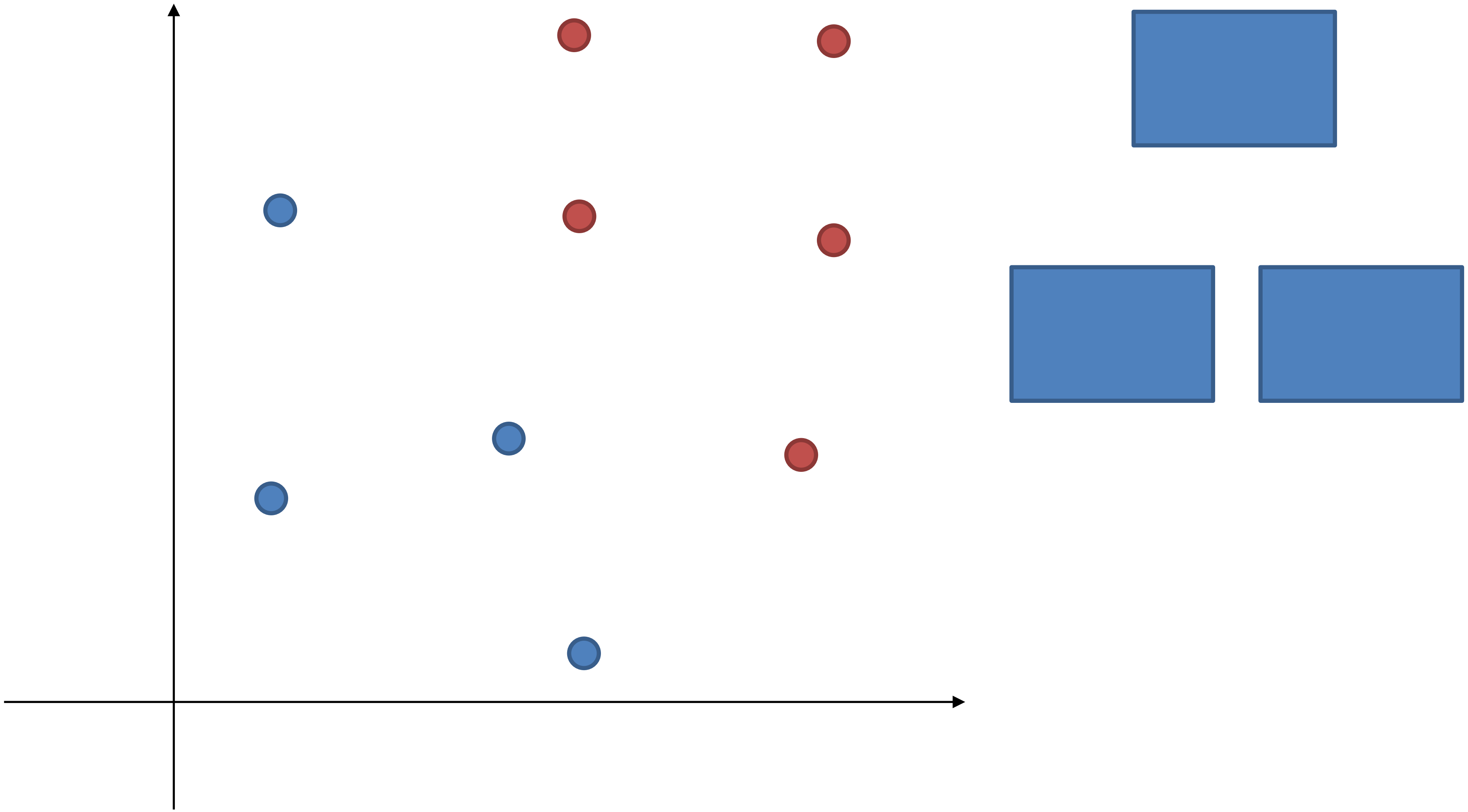
Malahanobis Distance

$$d(p,q)=\sqrt{(p-q)^T S^{-1}(p-q)}$$

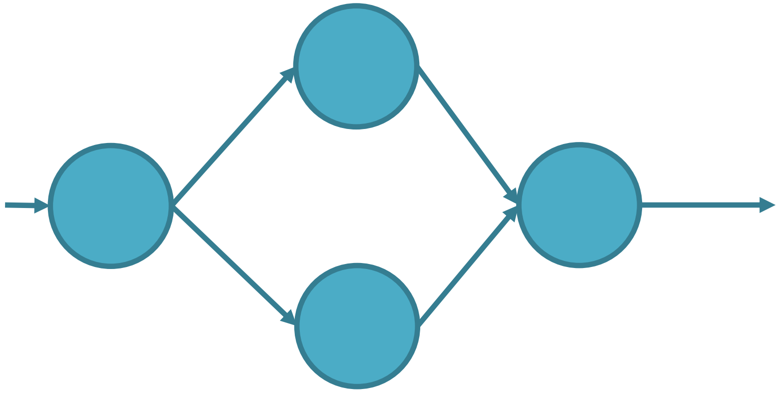
Support Vector Classifier

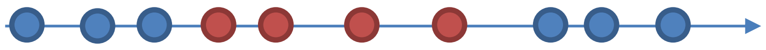


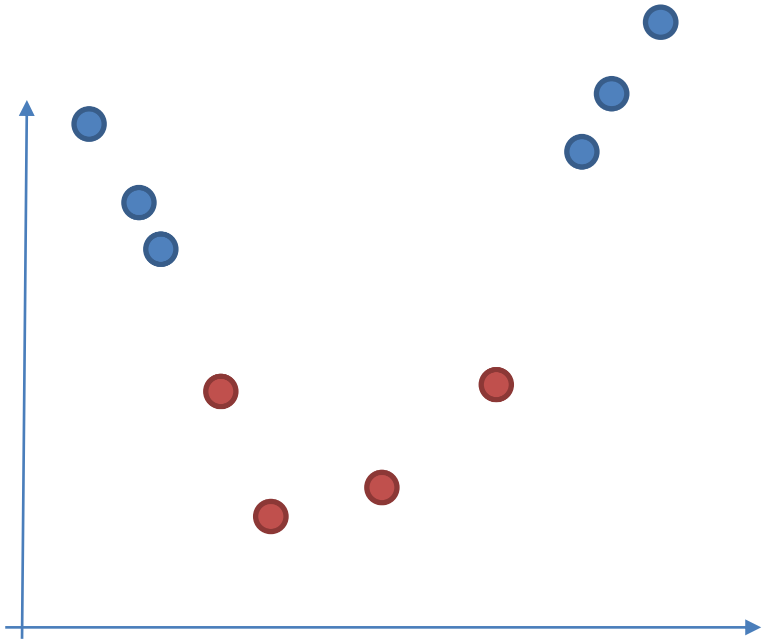
Decision Trees



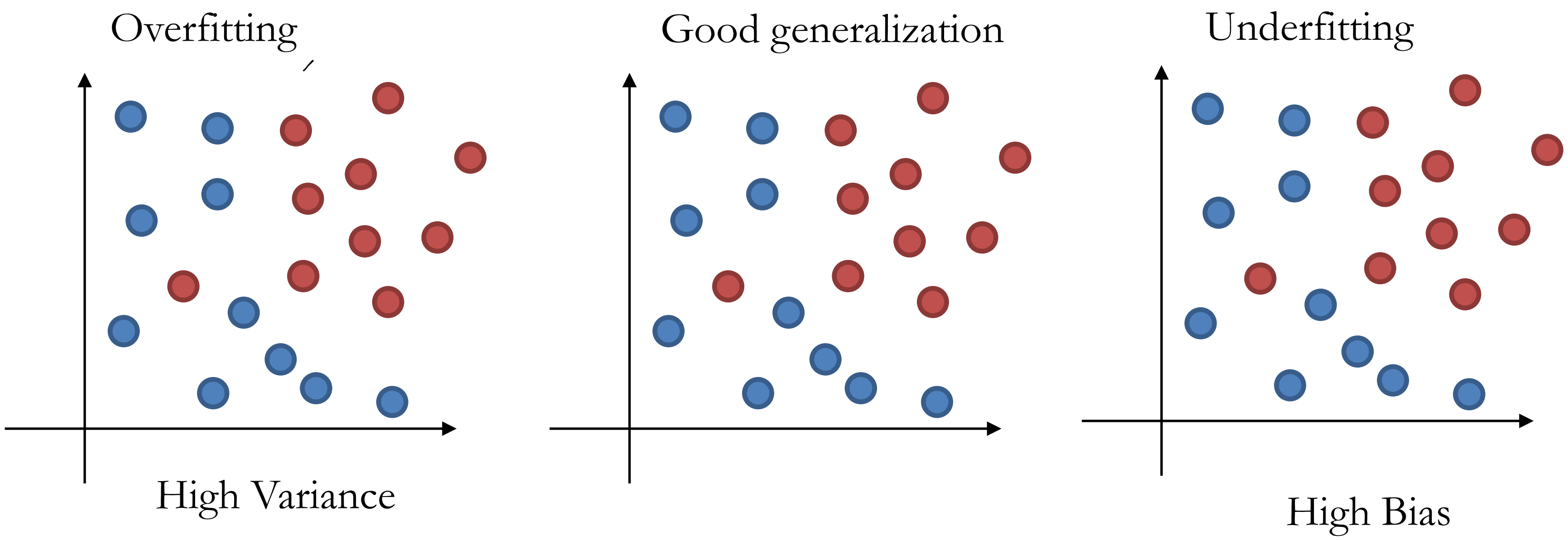
Multilayer Perceptron







Overfit and underfit in classification problems



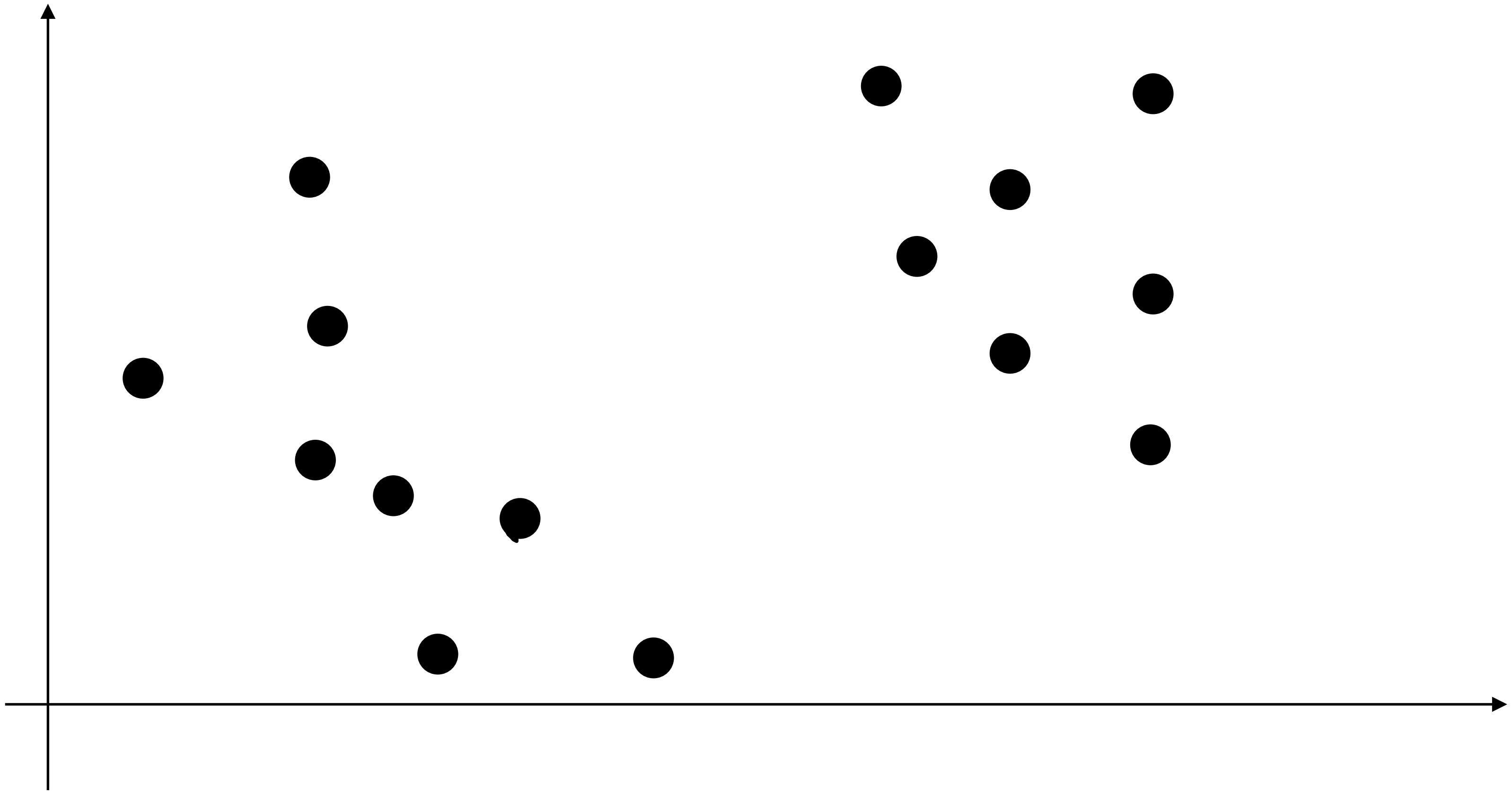
Carne en su jugo

Carne en su jugo (meat in its juice) is a typical dish from Guadalajara, Small pieces of flank steak are cooked in their juices, then mixed with whole beans and crispy crumbled bacon.

Let’s take a rest!

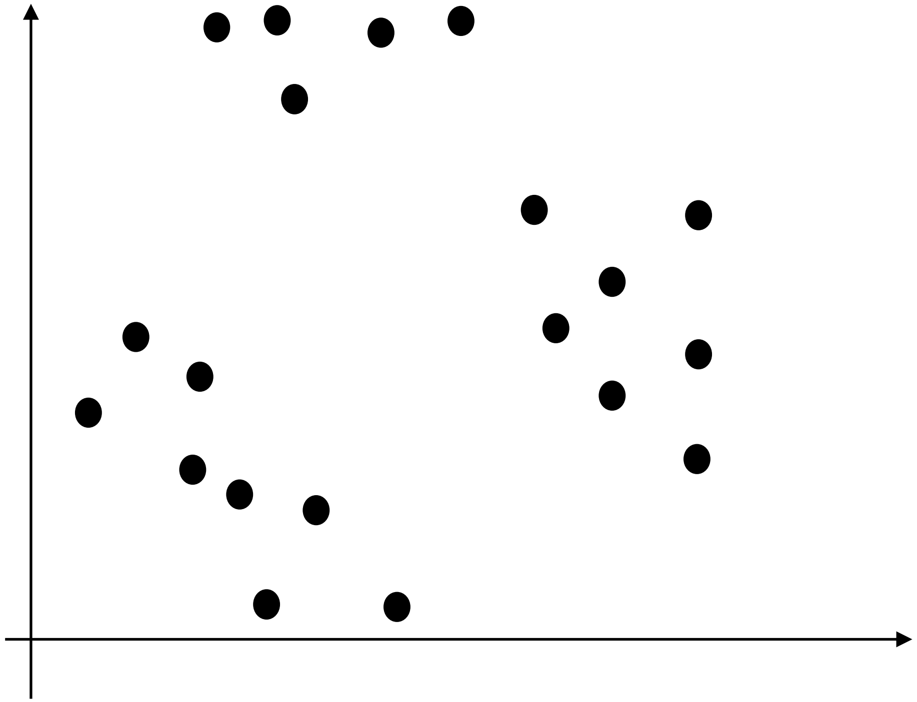
The Clustering Problem

- Clustering is an unsupervised problem with discrete output.



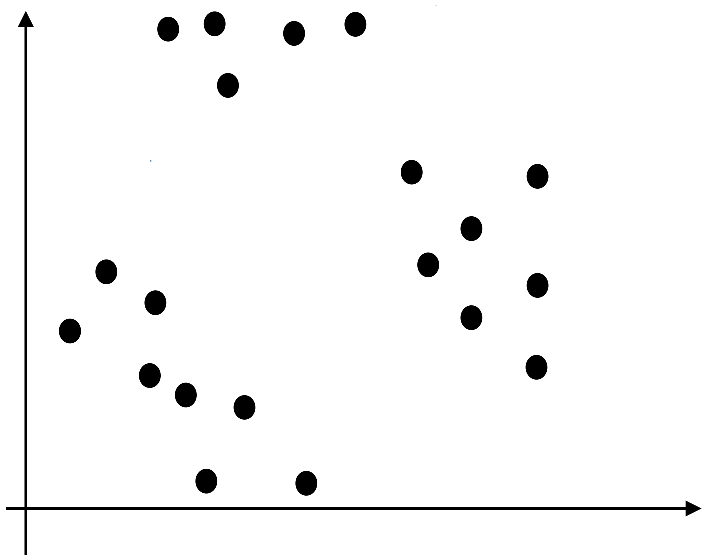
Types of clustering

- Partitional Clustering



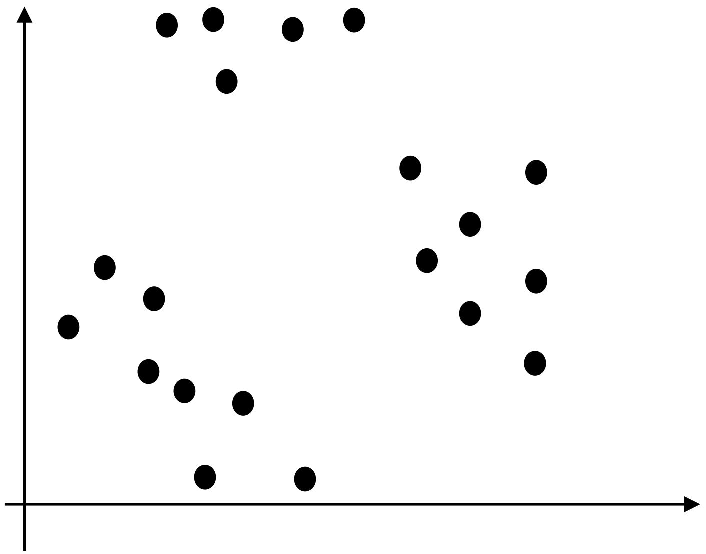
Types of clustering

- Hierarchical-based clustering



Bottom-Up

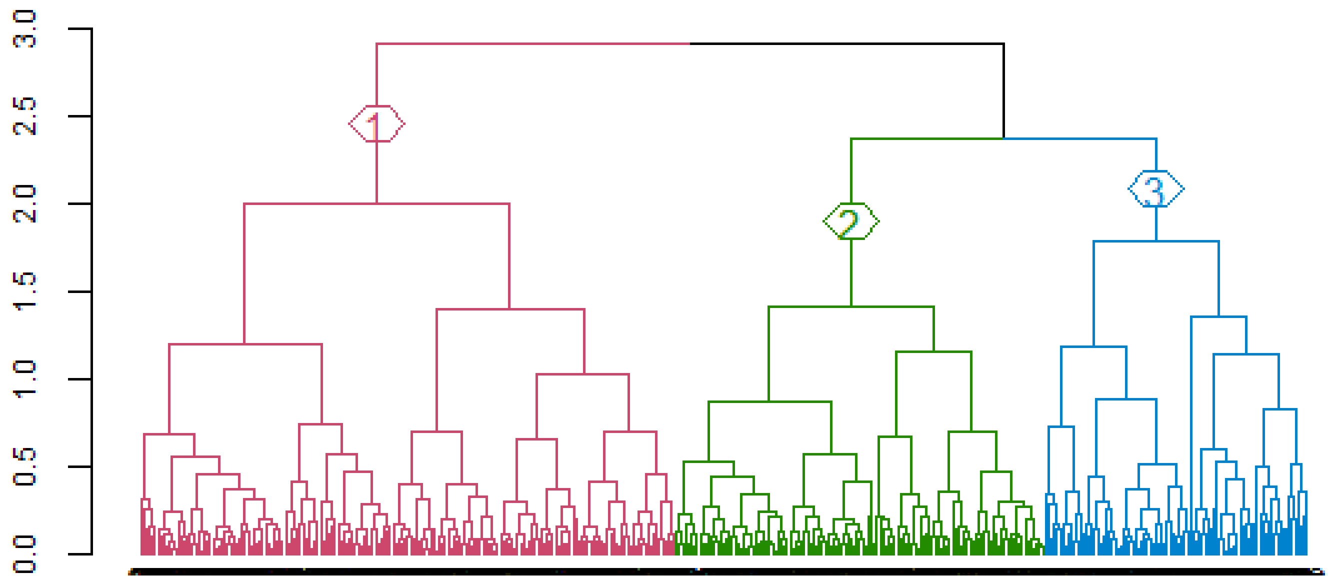
Agglomerative



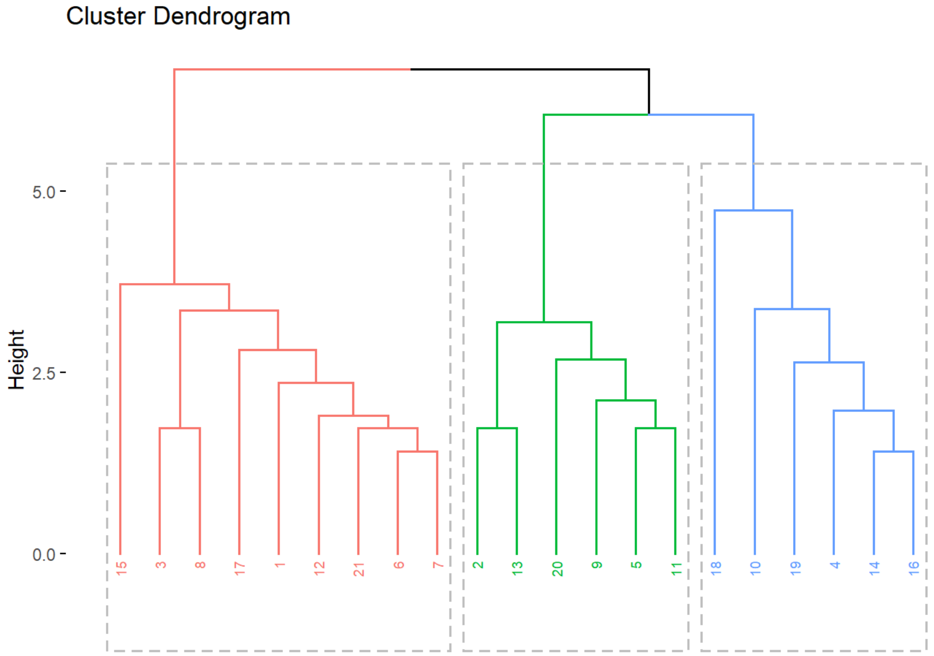
Top-Down

Divide and conquer

Dendrogram



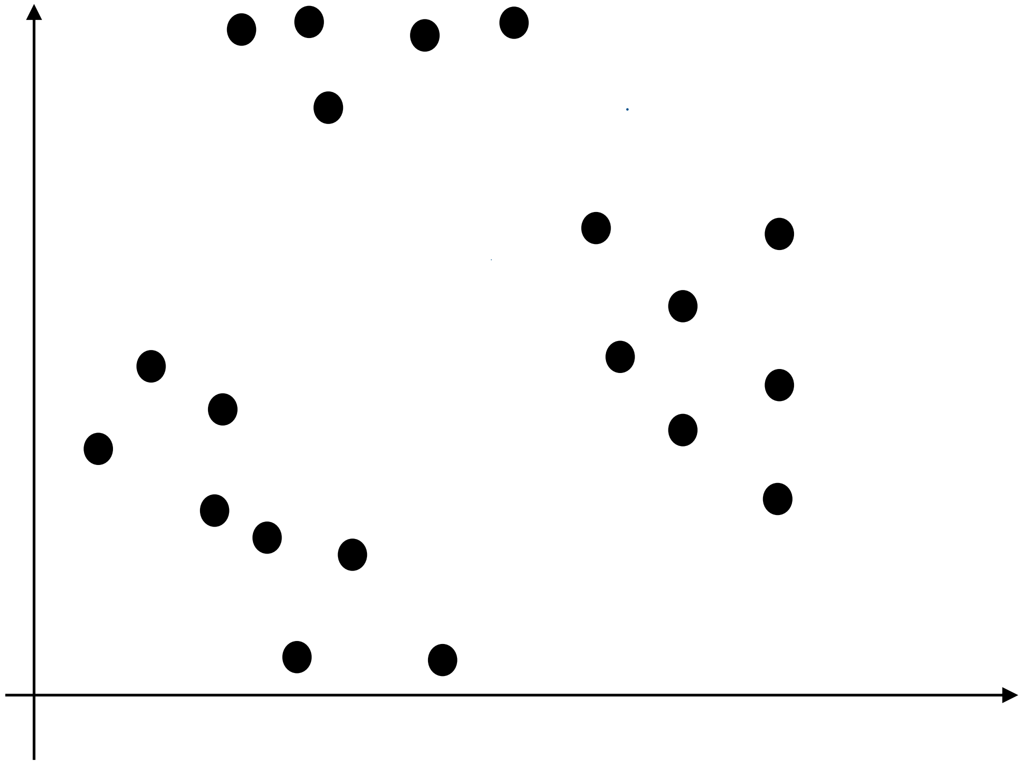
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3



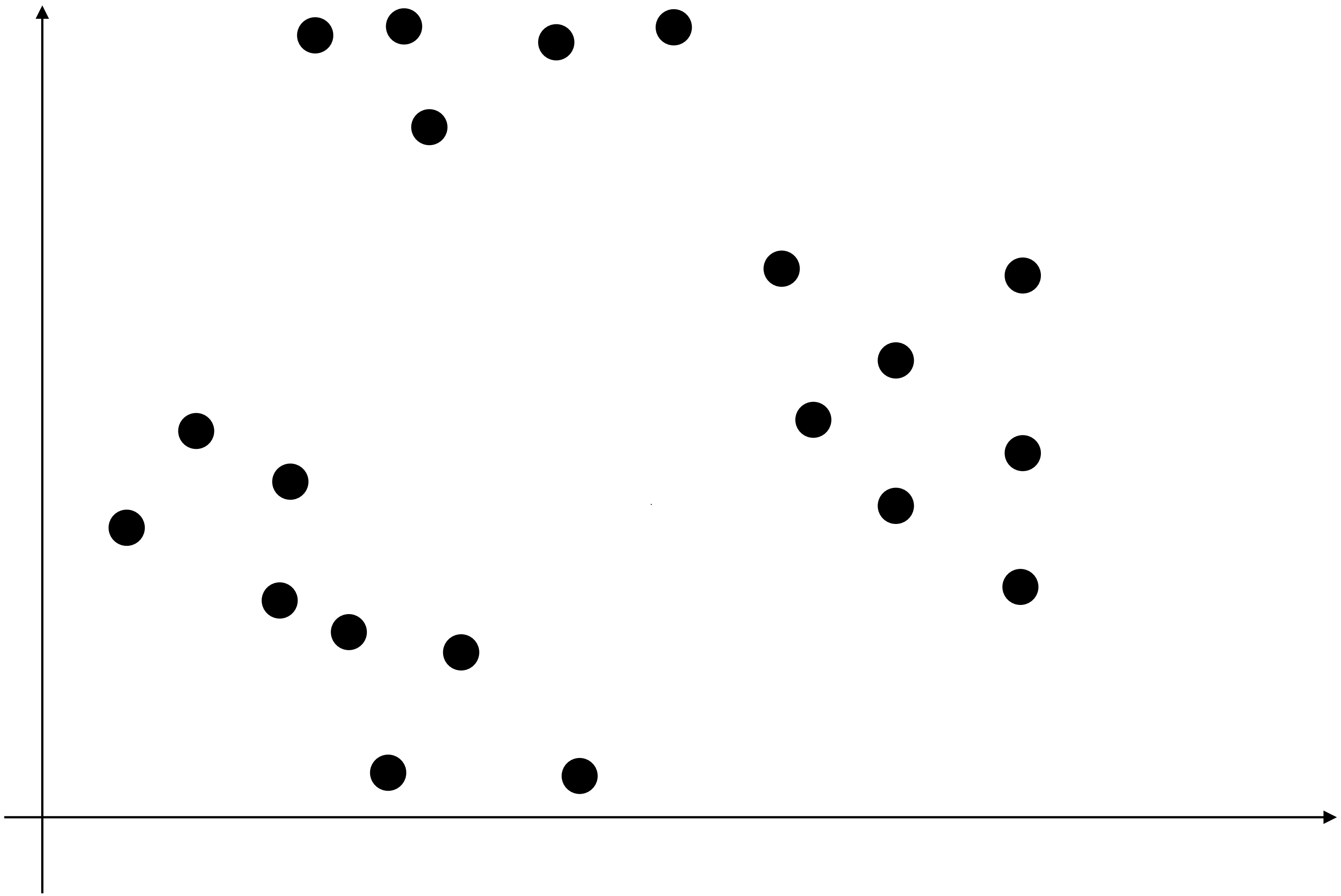
Cluster Dendrogram
Height
5.0
2.5
0.0
15
3
8
17
1
12
21
6
7
2
13
20
9
5
11
18
10
19
4
14
16

Types of clustering

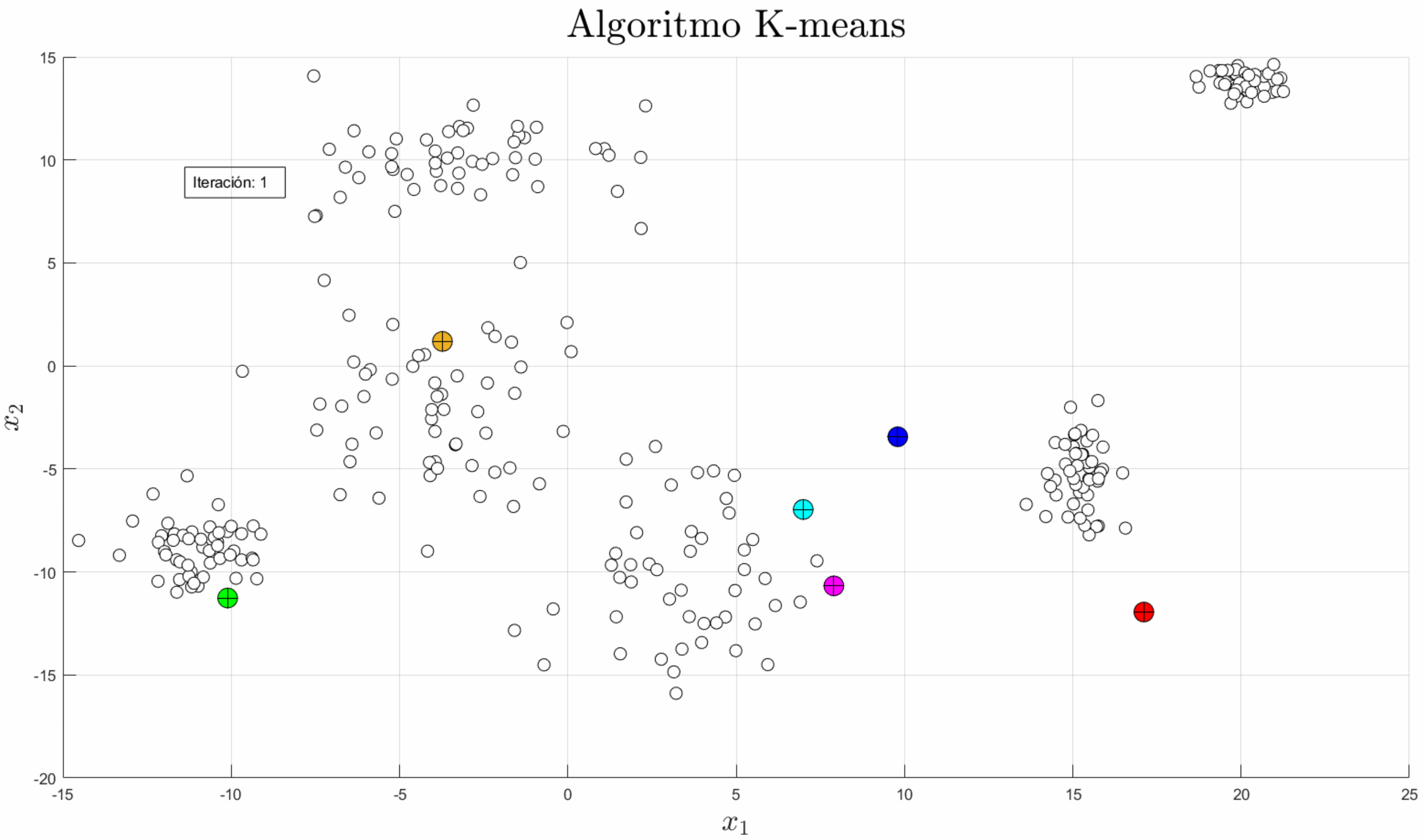
- Density-based clustering



K-means



K-means

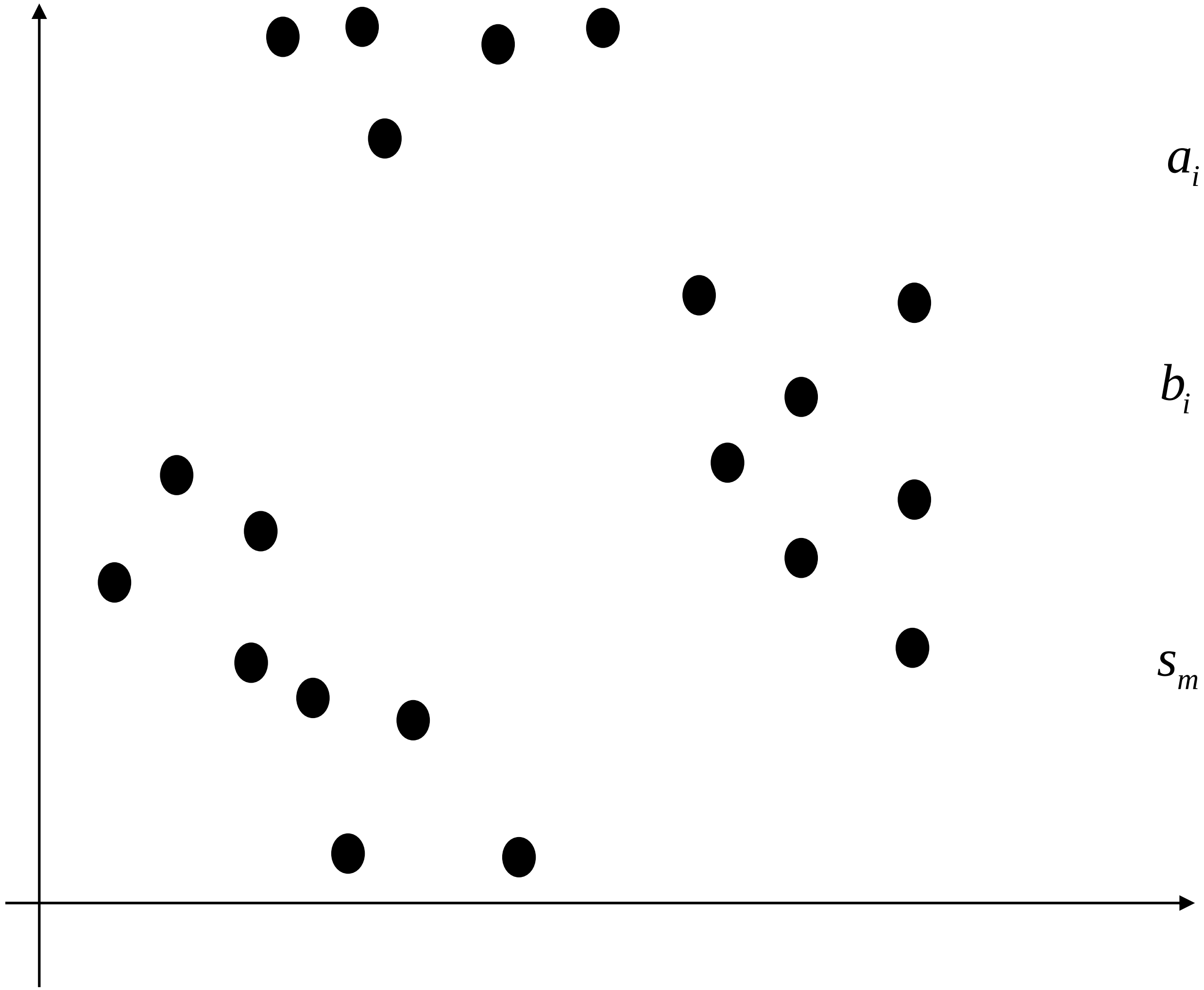


Algoritmo K-means
Iteración: 1
15
10
5
0
-5
-10
-15
-20
-15
-10
-5
0
5
10
15
20
25
x_2
x_1

K-means

- Variants
  - K-means++
  - K-means on-line
  - Mini-batch K-means
  - C-means

Silhouette Score



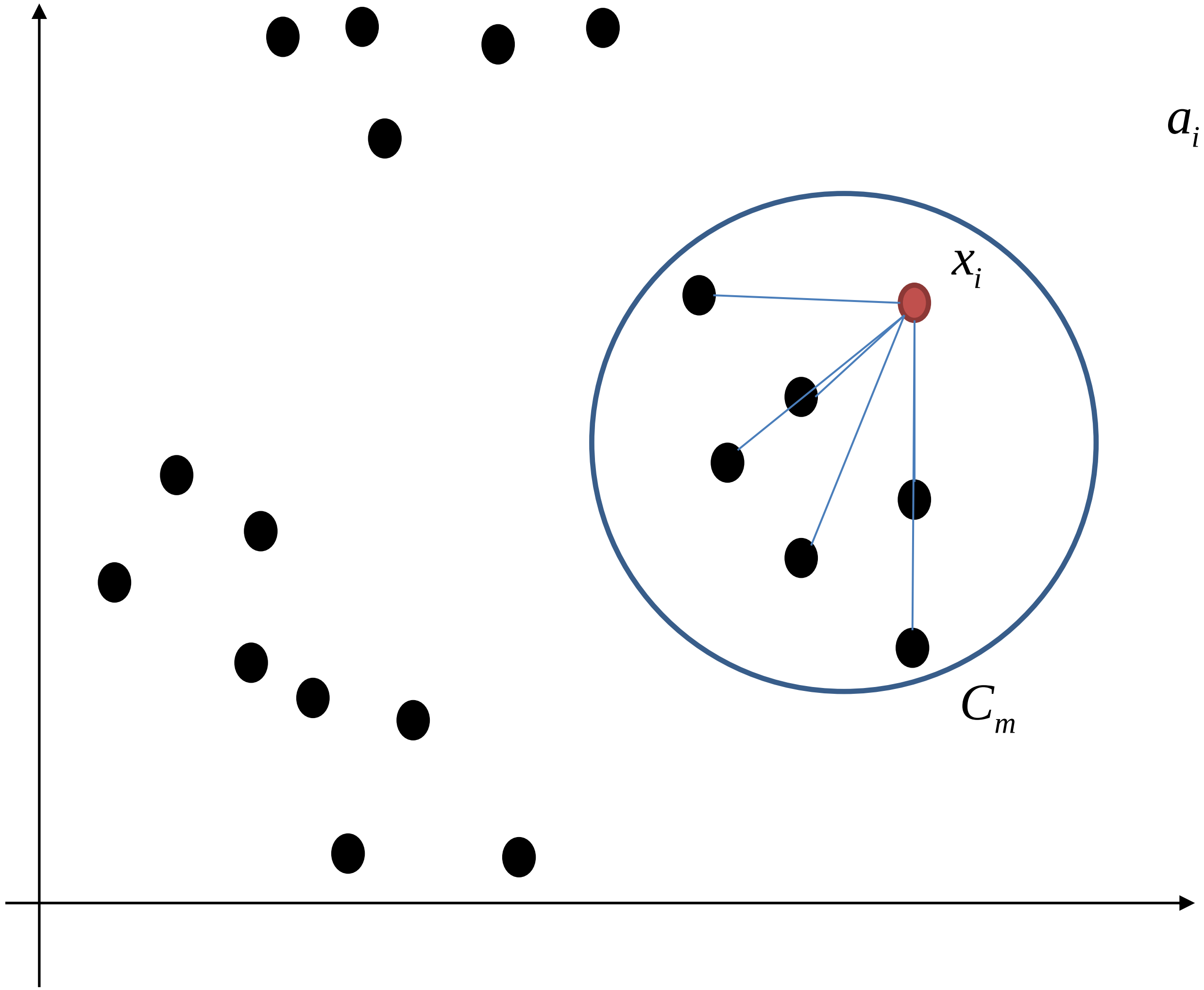
$$x_i \in C_m$$

$$a_i = \frac{1}{|C_m|-1} \sum_{x_j \in C_m} dist(x_i, x_j)$$

$$b_i = \min_{k \neq m} \frac{1}{|C_k|} \sum_{x_j \in C_k} dist(x_i, x_j)$$

$$s_m = \begin{cases} \dfrac{b_i - a_i}{\max\{a_i, b_i\}} & |C_i| > 1 \\ 0 & |C_i| = 1 \end{cases},$$

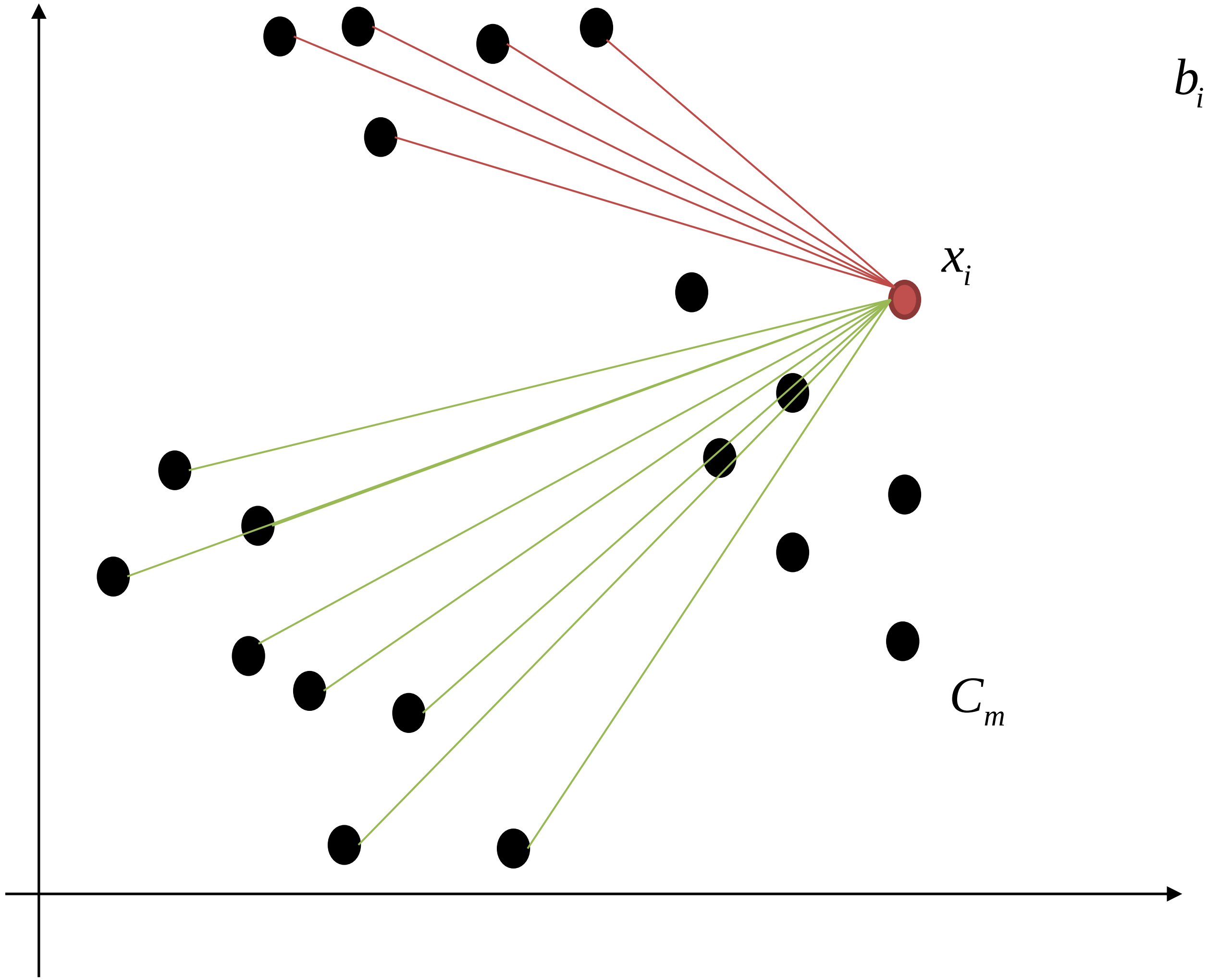
Silhouette Score



$$x_i \in C_m$$

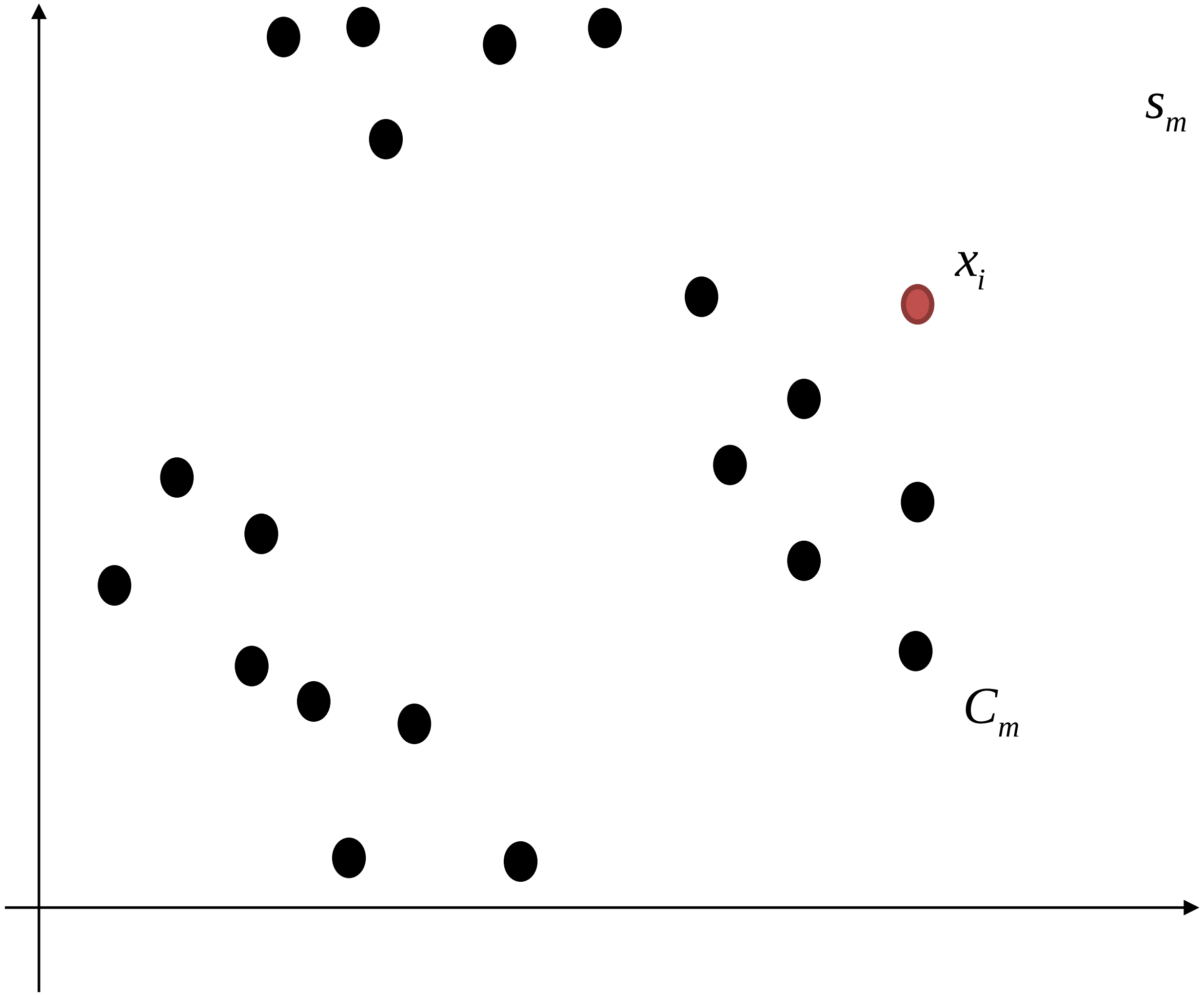
$$a_i = \frac{1}{|C_m| - 1} \sum_{x_j \in C_m} dist(x_i, x_j)$$

Silhouette Score



$$b_i = \min_{k \neq m} \frac{1}{|C_k|} \sum_{x_j \in C_k} dist(x_i, x_j)$$

Silhouette Score

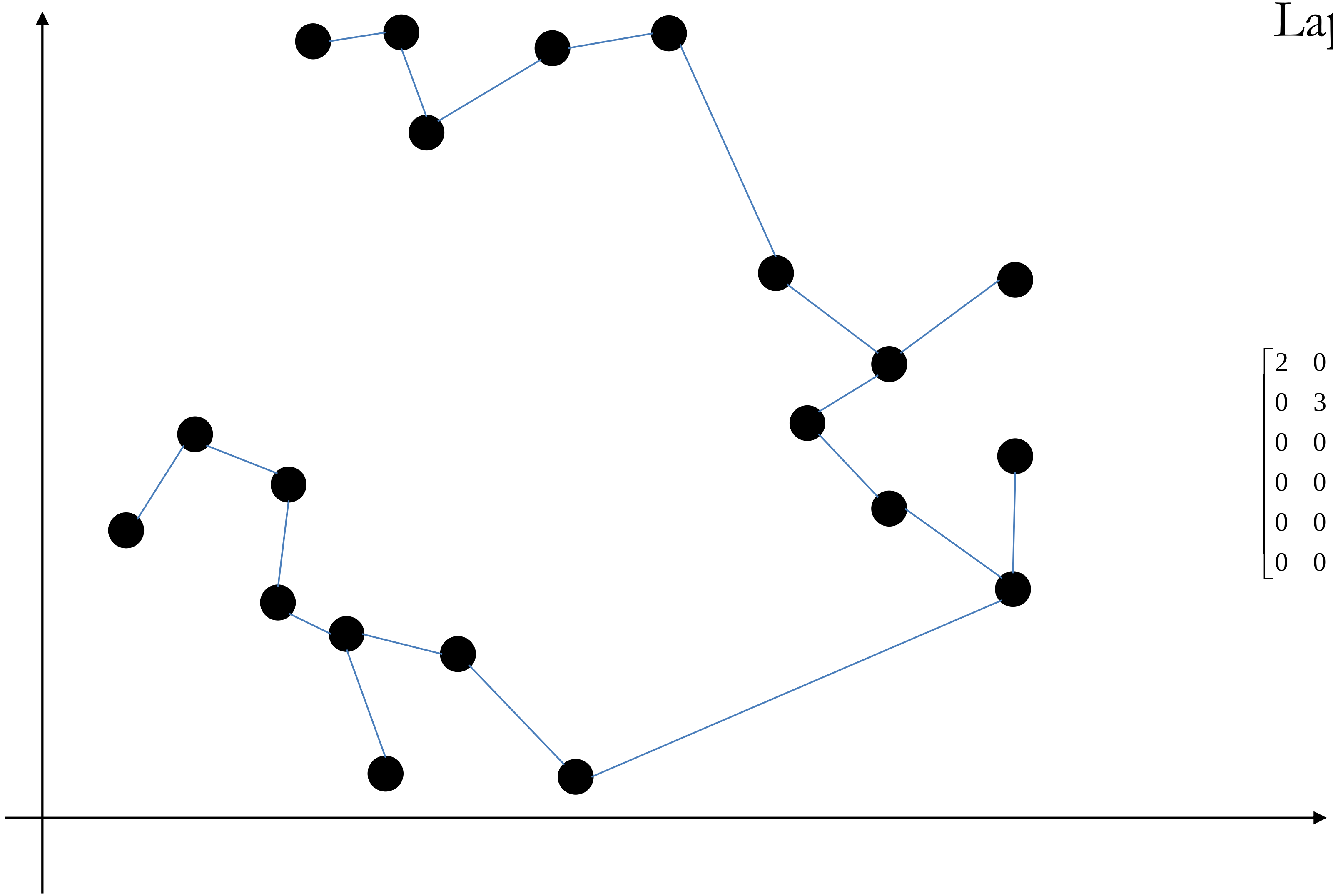


$$s_m = \begin{cases} \dfrac{b_i - a_i}{\max\{a_i, b_i\}} & |C_i| > 1 \\ 0 & |C_i| = 1 \end{cases},$$

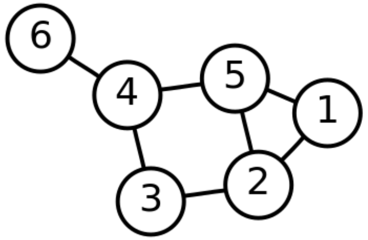
$$-1 < s_m < 1$$

$$s = \frac{1}{p} \sum_i s_i$$

Spectral Clustering



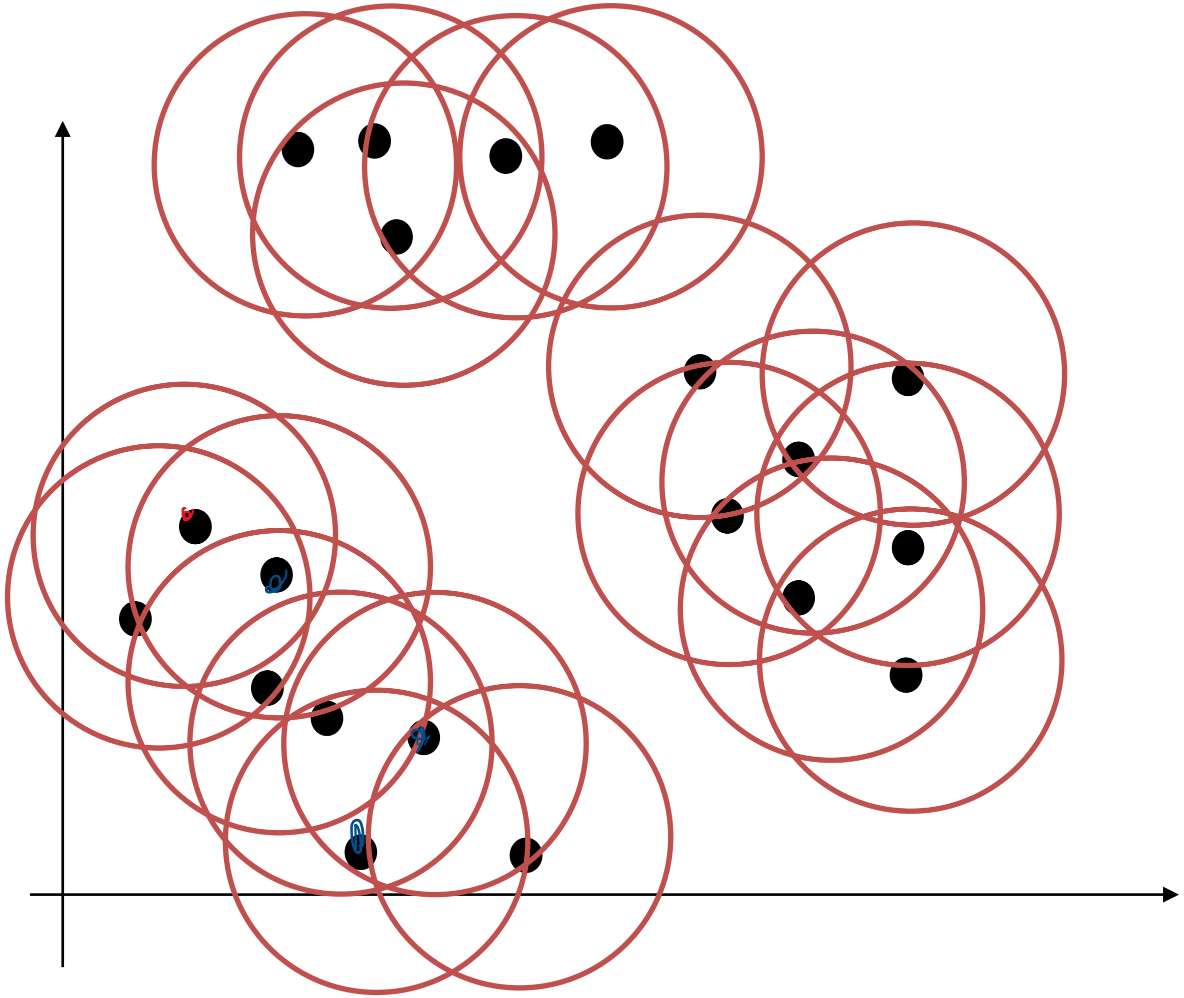
Laplacian Matrix of the Graph



$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

DBSCAN

- Density-Based Spatial Clustering of Application with Noise



Definition of dimensionality reduction

- In an unsupervised technique with continuous output

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \boxed{\text{Dimensionality reduction}} \longrightarrow y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$n > m$$

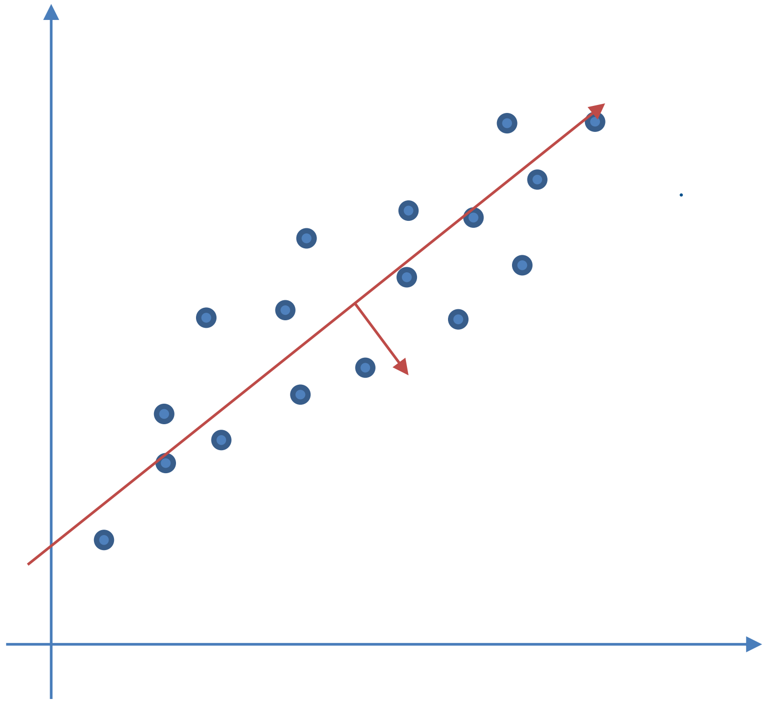
$$\left\{ x^{(1)}, x^{(2)}, \cdots, x^{(p)} \right\}$$

The Curse of Dimensionality

The concept of the curse of dimensionality was coined by Richard E. Bellman when he was working on dynamic programming problems.

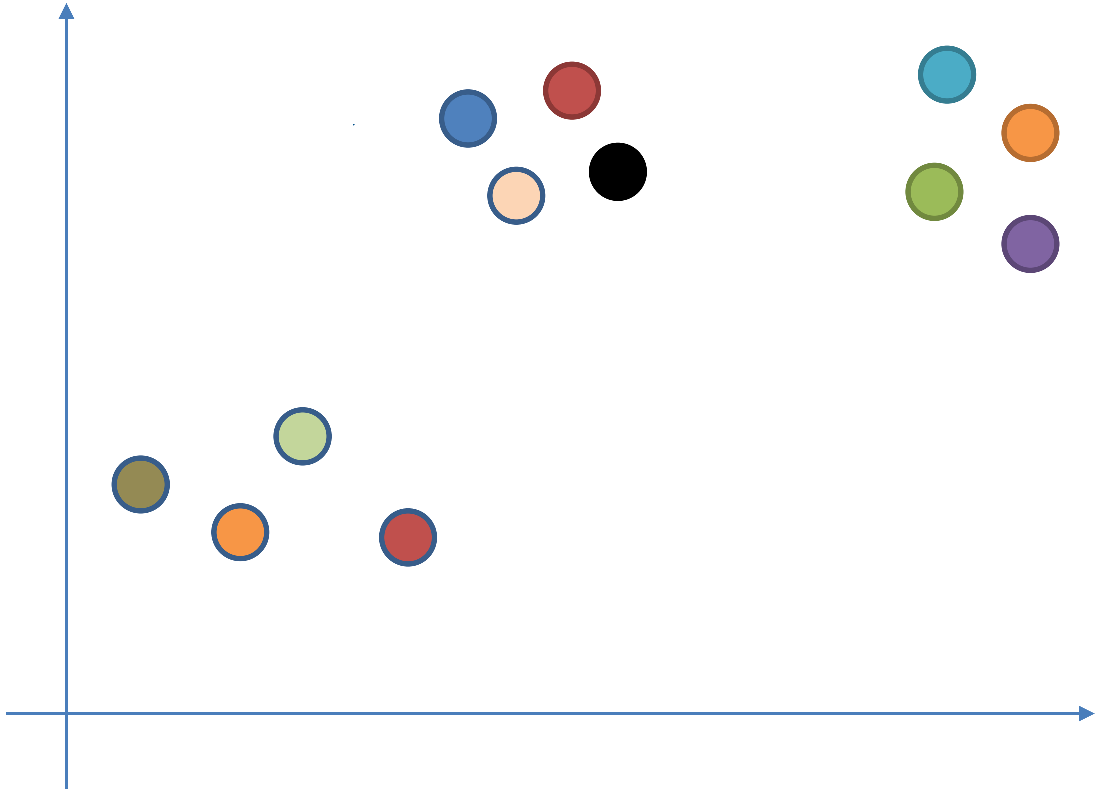
When we increase the dimensionality of a space, the volume of that space grows so fast that the data is scattered. This scattering of space causes problems in techniques that require statistical significance.

Principal Component Analysis(PCA)

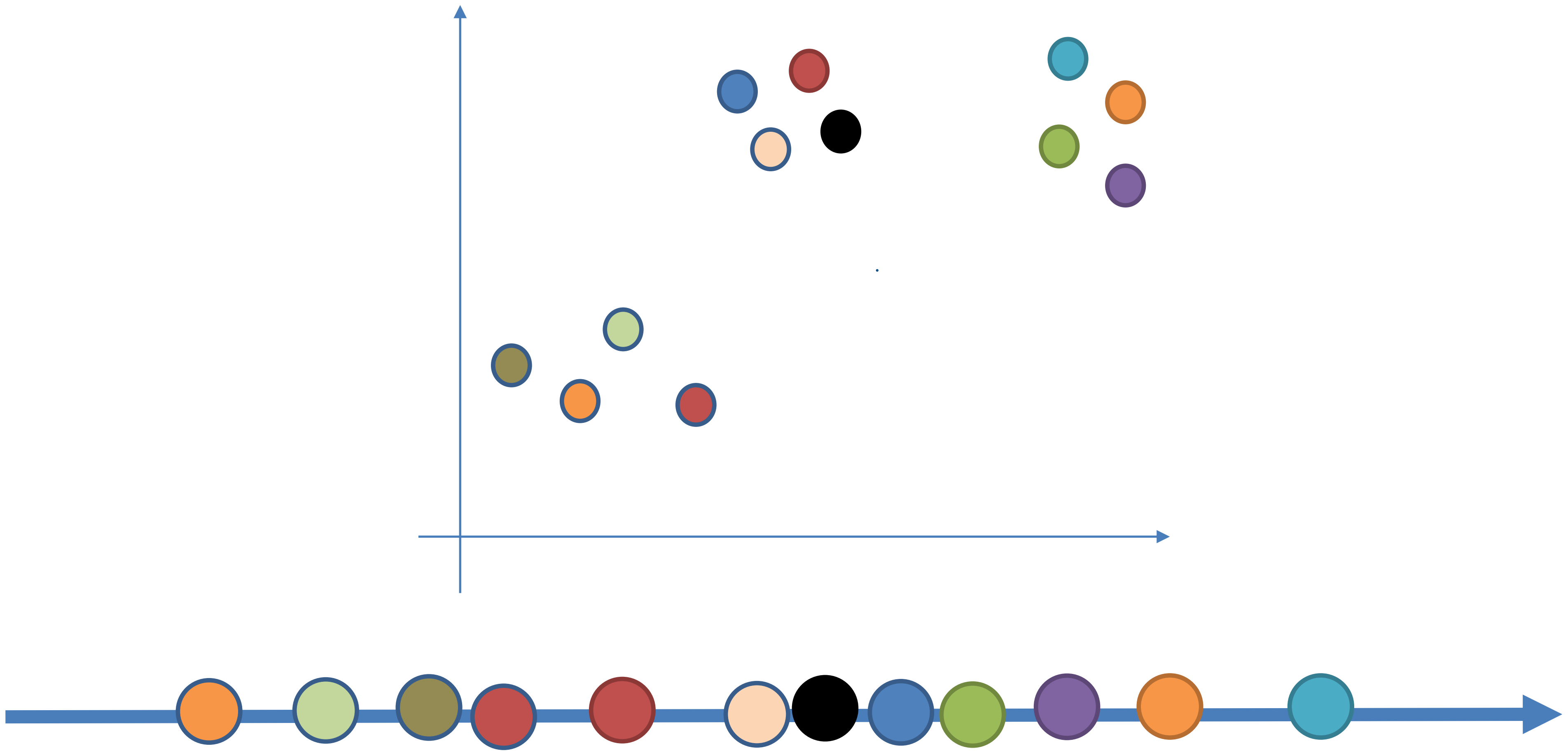


t-Stochastic Neighbors Embedding (t-SNE)

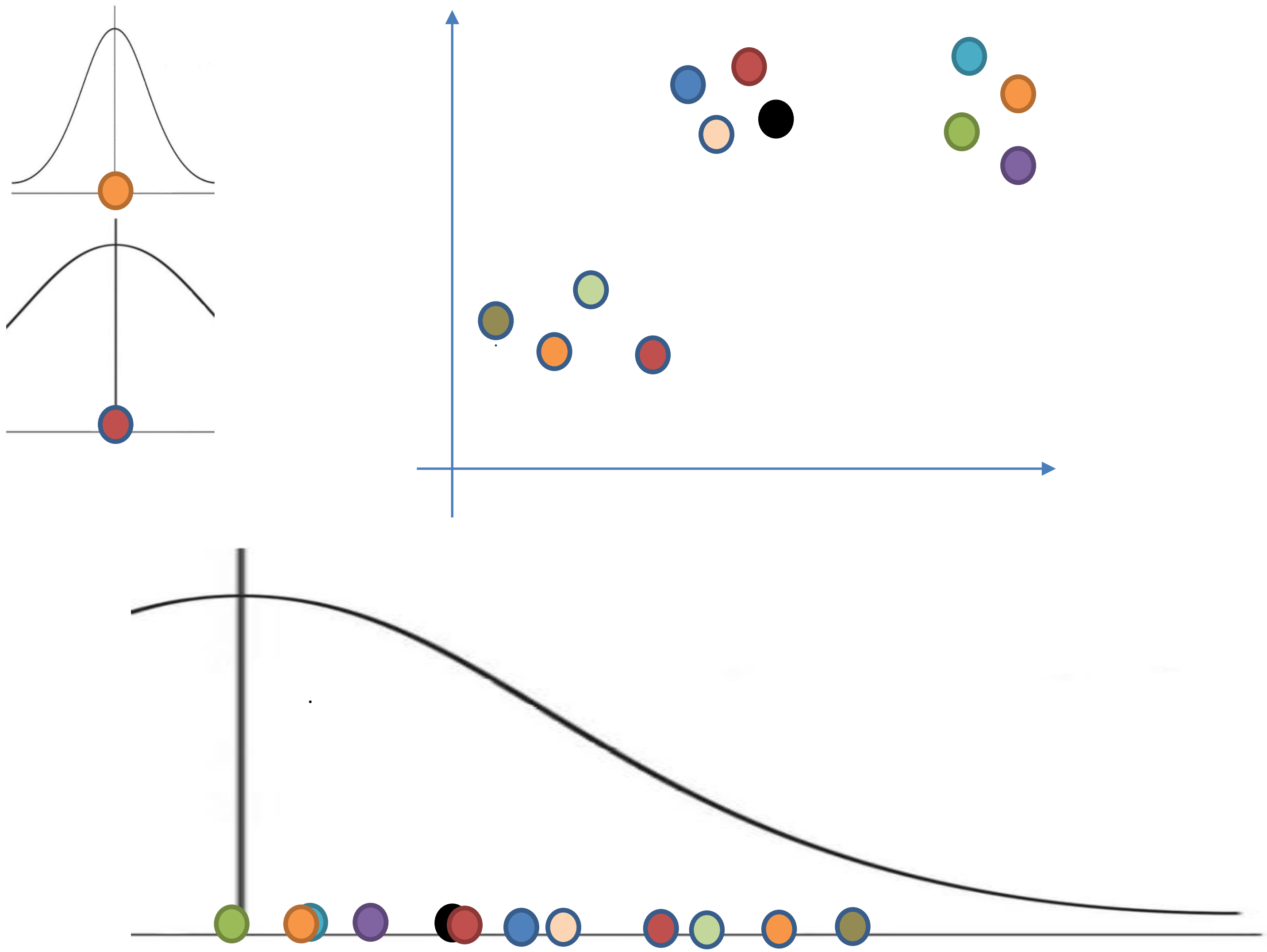
- It is a technique for reducing dimensionality to a very low dimension.



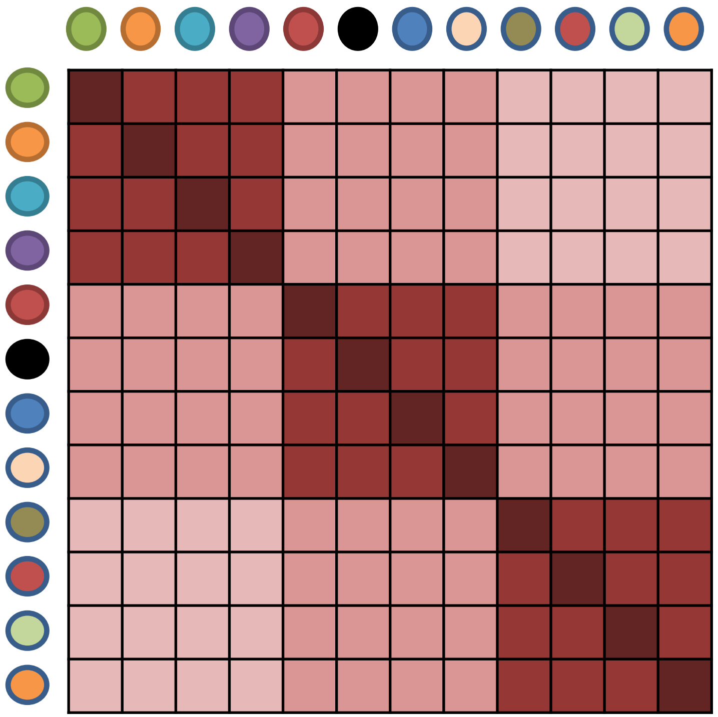
t-Stochastic Neighbors Embedding (t-SNE)



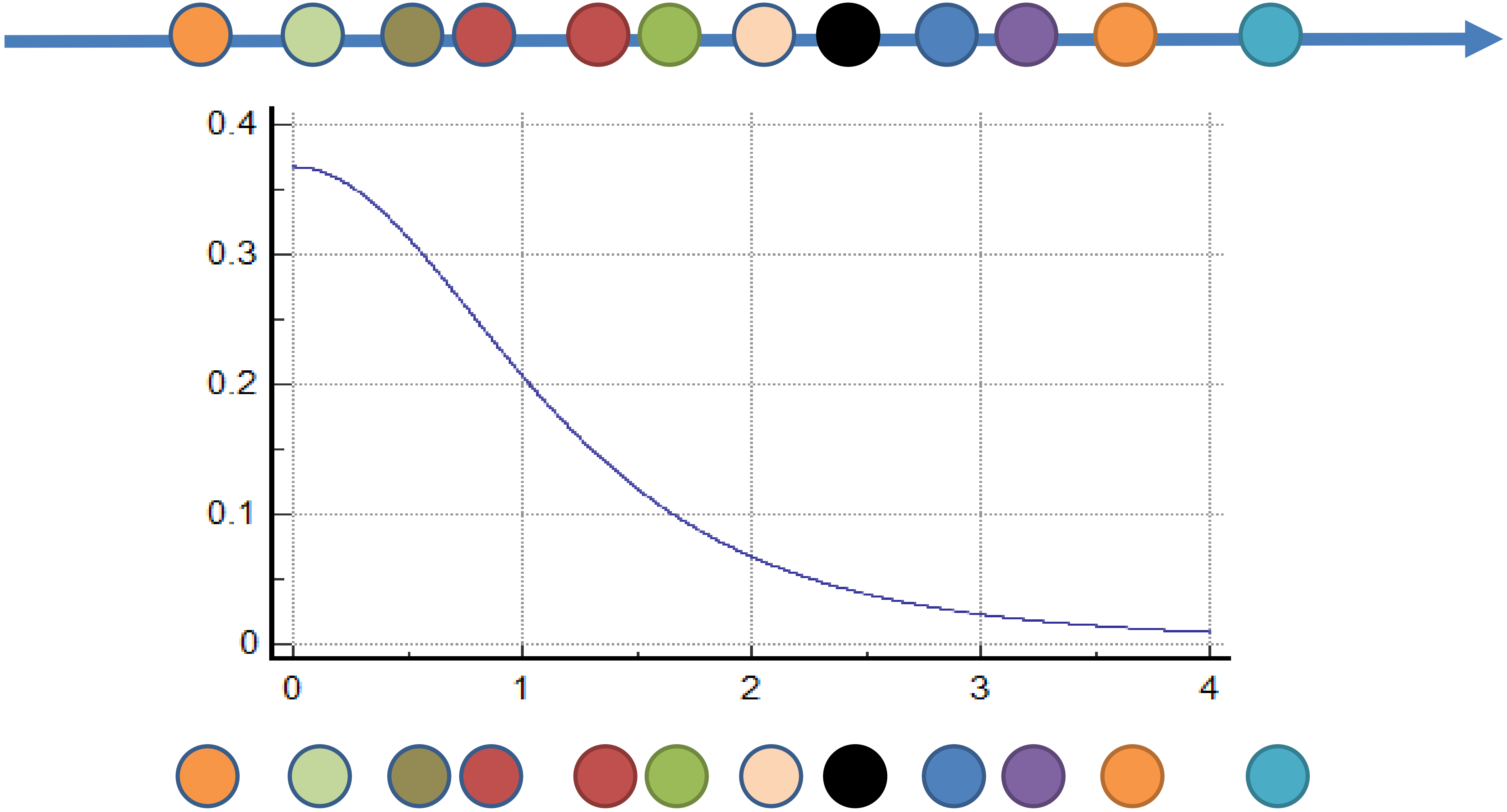
t-SNE



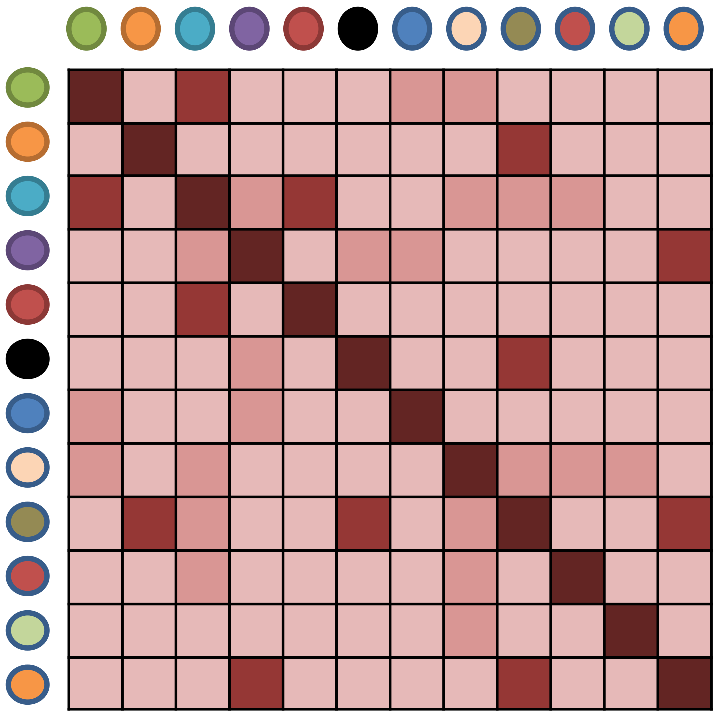
t-Stochastic Neighbors Embedding (t-SNE)



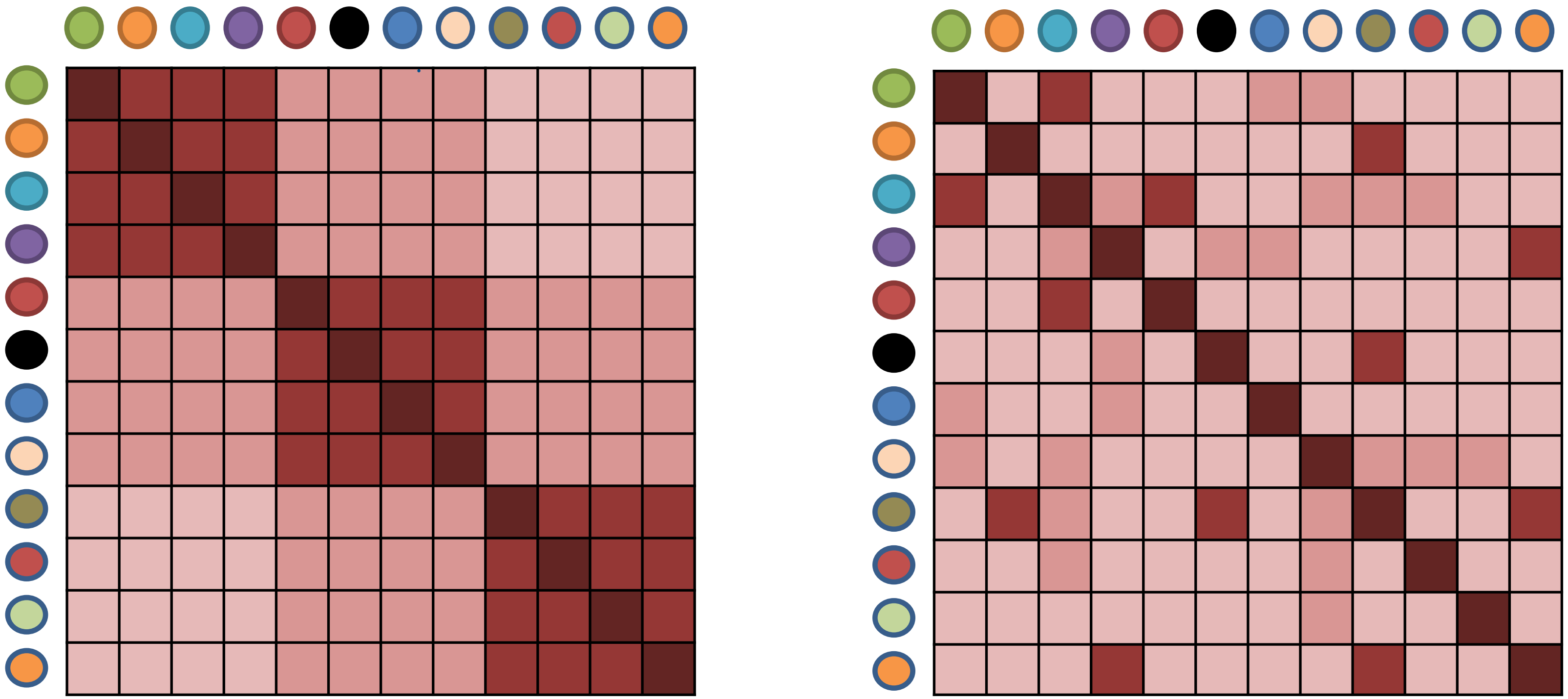
t-Stochastic Neighbors Embedding (t-SNE)



t-Stochastic Neighbors Embedding (t-SNE)



t-Stochastic Neighbors Embedding (t-SNE)



Creator Page: https://lvdmaaten.github.io/tsne/
It is advisable to review UMAP https://umap-learn.readthedocs.io/en/latest/
Tensorflow projector https://projector.tensorflow.org/