Topological Data Analysis and Persistence Theory

NSF/CBMS Conference

Valdosta State University

August 8-12, 2022

Peter Bubenik, University of Florida

Lecture 6: TDA and Machine Learning

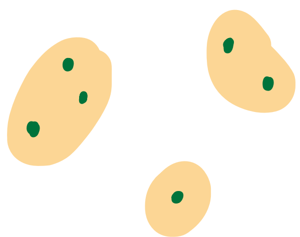
Outline:
1. Clustering
2. Classification
3. Regression
4. Deep Learning

Please interrupt me!!!

1. Clustering

1.1 The Learning Problem

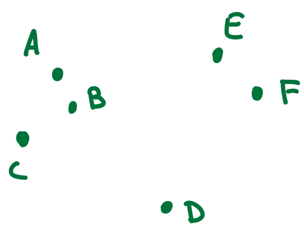
Given $x_1, \dots, x_N \in \mathbb{R}^d$ or more generally $X = \{x_1, \dots, x_N\}$ with a metric d partition X into components called clusters.

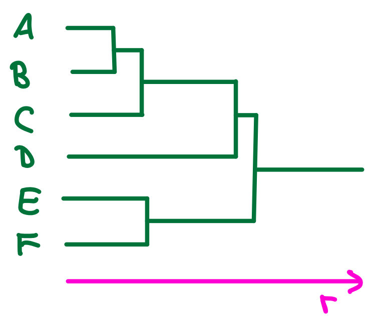


Unsupervised Learning

1.2 Single-linkage clustering

Hierarchical clustering : increasingly merge elements of X





Merge Tree

Single linkage clustering : merge clusters according to the smallest distance between any of their elements

1.3 Persistent π_0

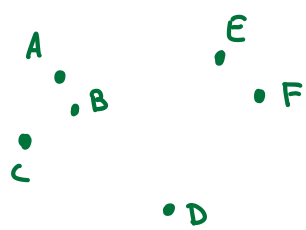
$\pi_0 : \mathbf{Top} \to \mathbf{Set}$ or $\pi_0 : \mathbf{Simp} \to \mathbf{Set}$

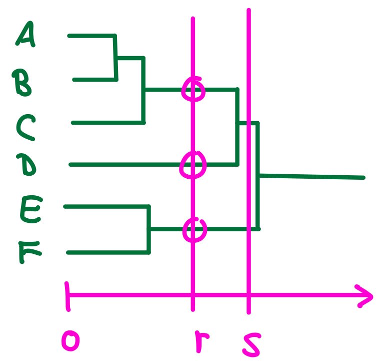
sends a space or complex to its set of path components.

Given (X,d) above $\quad VR\,X : \mathbb{R} \to \mathbf{Simp}$

Consider $\quad \mathbb{R} \xrightarrow{VR\,X} \mathbf{Simp} \xrightarrow{\pi_0} \mathbf{Set}$

<u>Lemma</u> $\pi_0\, VR\, X$ is isomorphic to the merge tree in single linkage clustering.

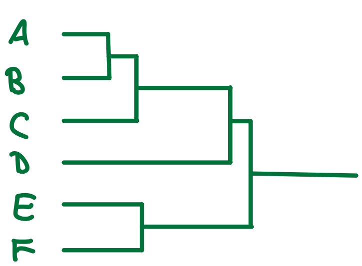




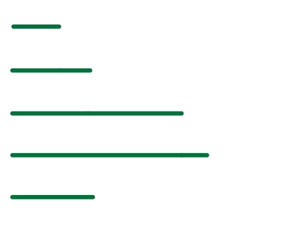
$$\pi_0\, VR_r X = \{ \{A,B,C\}, \{D\}, \{E,F\} \}$$

$$\pi_0\, VR_s X = \{ \{A,B,C,D\}, \{E,F\} \}$$

We can obtain the barcode for persistent H_0 from the merge tree for persistent π_0 by forgetting the mergers.



$\longrightarrow$



Advise: If you are using TDA and only using H_0 then you may be better off using a more sophisticated version of clustering.

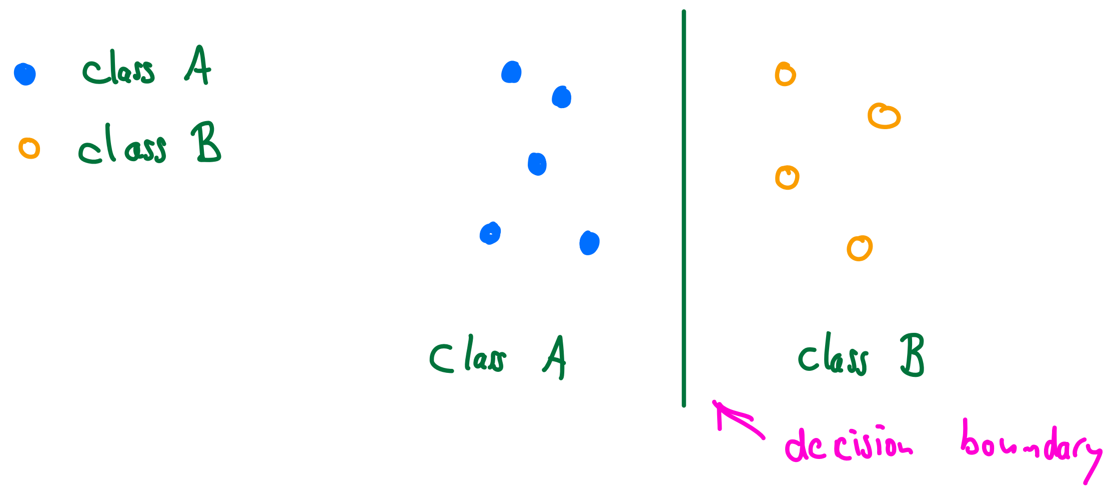
2. Classification

2.1 The Learning Problem

Supervised Learning

Given $x_1, \ldots x_n \in \mathbb{R}^d$ in class A and
$y_1, \ldots, y_m \in \mathbb{R}^d$ in class B

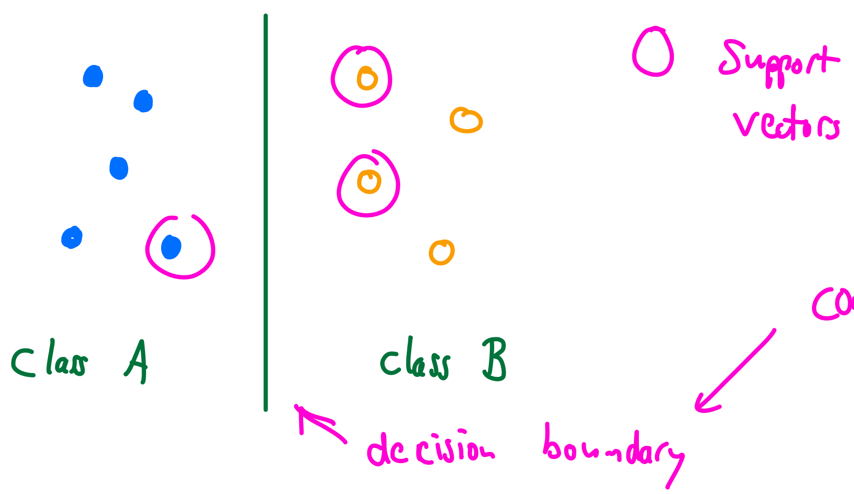
we want to give a function $\mathbb{R}^d \to \{A, B\}$ that accurately classifies unlabeled data. called a classifier



2.2 Support Vector Machine Classification

(Hard Margin)

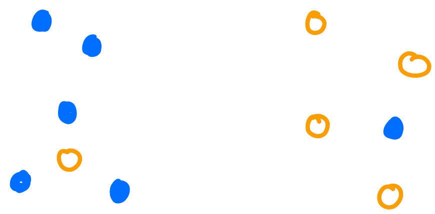
Support Vector Machine (SVM) finds the separating hyperplane that maximizes the Hausdorff distance to the two classes.



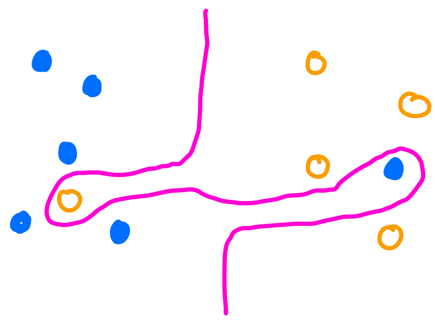
compute via least squares

fast!

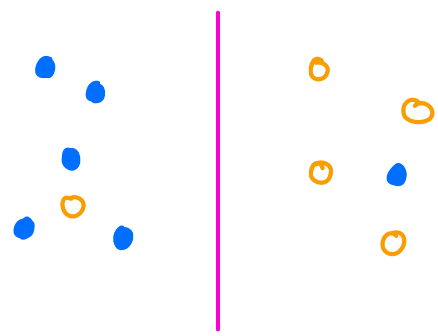
What if there is no separating hyperplane?



Which decision boundary do we prefer?



or



(Soft Margin)

SVM finds optimal hyperplane where points on the wrong side are weighted by some "cost".

"hyperparameter" eg Cost = 10

2.3 Assessing a classifier

We have data $x_1, \dots x_n \in \mathbb{R}^d$ in class A

$y_1, \dots y_m \in \mathbb{R}^d$ in class B

and an algorithm for building a classifier.

If we use all of our data to build a classifier we cannot assess its accuracy on new data.

Idea: split our data into <u>training data</u> used to build a classifier and <u>testing data</u> used to determine classification accuracy.

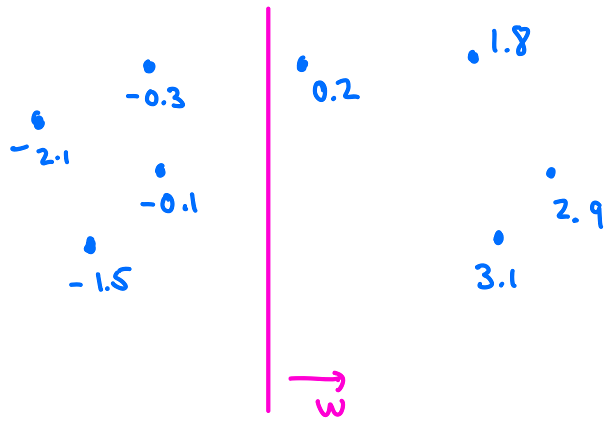
3 Regression

3.1 The Learning Problem

We are given $x_1, \dots x_n \in \mathbb{R}^d$ together with values $y_1, \dots y_n \in \mathbb{R}$

We want to learn $f: \mathbb{R}^d \to \mathbb{R}$ that predicts $y \in \mathbb{R}$ given $x \in \mathbb{R}^d$.

3.2 Support Vector Machine Regression



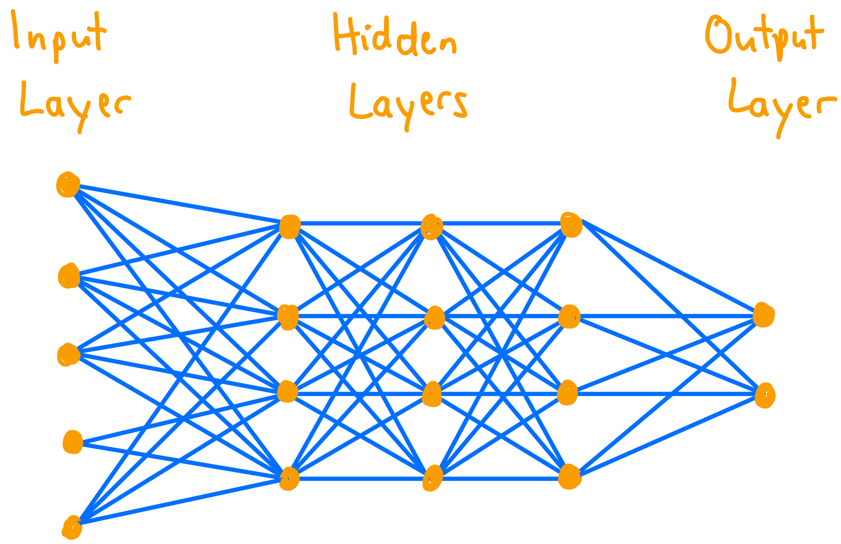
Idea: Find a hyperplane and normal vector so that f is given by the signed (weighted) distance to the hyperplane.

$$y = \langle w, x \rangle + b$$

4. Deep Learning

4.1 MultiLayer Perceptron (MLP)

Example of Deep Neural Network



Idea: • each edge has a weight

• inputs to a node are scaled by the weights and added and then composed with a nonlinear function

• on training the output is assessed by a loss function

• the error is backpropagated using partial derivatives

• the weights are adjusted

4.2 Topological Loss Function

Idea: TDA can be used to give a loss function

4.3 Topological Layer

Idea: TDA can be used to give a layer for the MLP.

Example: PLLay uses the persistence landscape as a layer in a Neural Network.