Topological Data Analysis and Shape Recognition

Ra-Zakee Muhammad
Pomona College

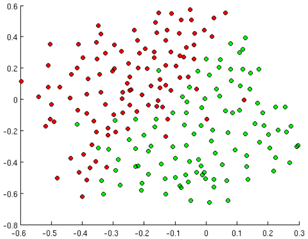
June 19, 2022

Table of Contents

The classification problem

Two Class Data Classification

Data is of the form $(\mathbf{x}, y)$ where $\mathbf{x}$ may be a vector of features, and $y \in \{0, 1\}$, known as the *class*. Given training data, the goal is to find a decision boundary for classifying future observations where only $\mathbf{x}$ will be available.



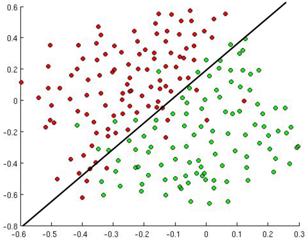
(Source: OpenClassroom Stanford)

Support Vector Machines (SVM) is a method of finding optimal linear boundaries.

The classification problem

Two class Data Classification

Data is of the form $(\mathbf{x}, y)$ where $\mathbf{x}$ may be a vector of features, and $y \in \{0, 1\}$, known as the *class*. Given training data, the goal is to find a decision boundary for classifying future observations where only $\mathbf{x}$ will be available.



(Source: OpenClassroom Stanford)

Support Vector Machines (SVM) is a method of finding optimal linear boundaries.

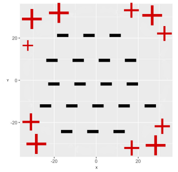
Non-Linear Boundaries

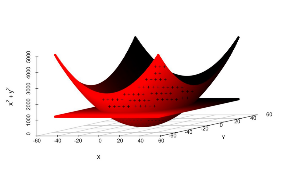
When lines don't work?

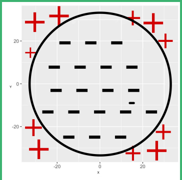
Mapping data into a feature space $\phi : \mathbb{R}^d \to \mathbb{R}^D$ induces non-linear decision boundaries in $\mathbb{R}^d$.

$\phi : \mathbb{R}^2 \to \mathbb{R}^3$

Consider 2 dimensional data that is mapped into $\mathbb{R}^3$ by adding the feature $z = x^2 + y^2$. We denote the two classes by (+) and (-).





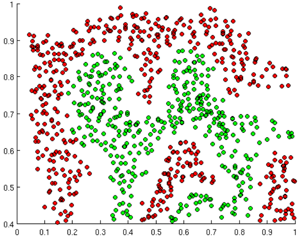


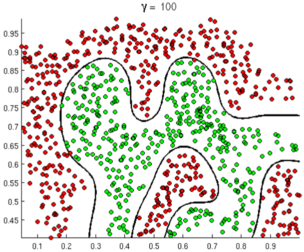
(Source: Rstudio)

The Kernel "Trick"

Note

The algorithm doesn't actually require these maps, only their inner products $<\phi(\mathbf{x}), \phi(\mathbf{y})>$. A kernel function is a function $K(\cdot, \cdot)$ such that $K(x, y) =< \phi(\mathbf{x}), \phi(\mathbf{y})>$. For instance the radial basis function $K(x, y) = e^{-\gamma||x-y||^2}$ corresponds to $\phi(x) \in \mathbb{R}^{\infty}$.





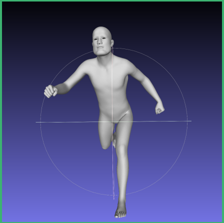
(Source: OpenClassroom Stanford)

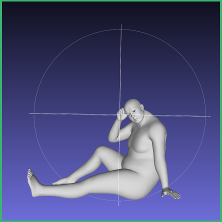
Classifying Non-Euclidean Object Data

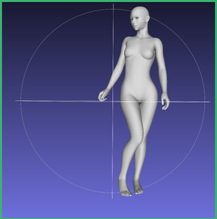
Introducing Object Data - Shape Retrieval

Our data set consists of 20 different poses for each of 15 bodies. The goal is, given a new pose of an existing body, to identify the correct class (body). (Data source: SHREC 2014 NIST)

Example Observations



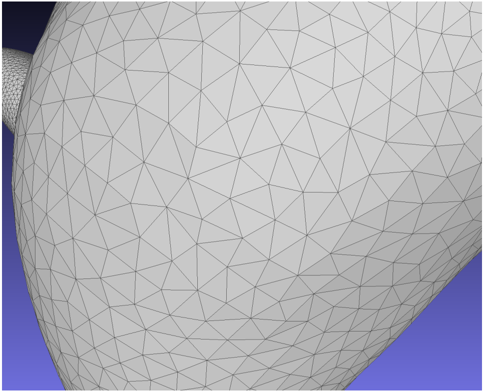




(Source: MeshLab)

First, what are you looking at?

Zooming into the shoulder of the first figure, we see this image.



(Source: MeshLab)

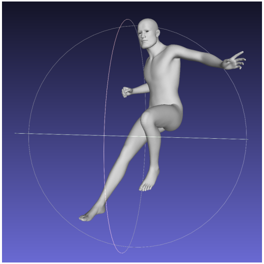
Simplicies

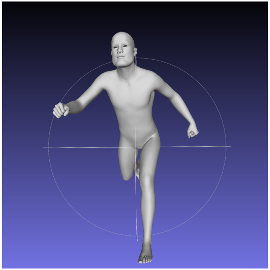
These triangles, edges and vertices are called (2, 1 and 0 dimensional) simplicies and collectively form a simplicial complex. They approximate the body surface manifold.

Procedural Detail

How do we consolidate curvature information?

Coordinates are insufficient for the classification problem. Curvature should be roughly maintained across poses.





(Source: MeshLab)

Heat Diffusion to the Heat Kernel

Heat Kernel

The Heat Kernel $K_t(x, y)$ has the property that for a Brownian Motion starting at point x at time, the probability that it will be in a set C at time t is $P(X_t^x \in C) = \int_C K_t(x, y)dy$.

Curvature

$K_t(x, y)$ is constructed from the eigen-pairs of Laplace-Beltrami operator, a manifold version of the Laplacian $\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$.

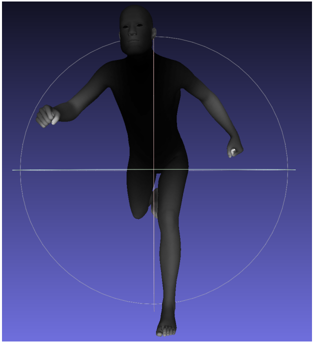
Approximation

On simplices, this is approximated by the graph-Laplacian.

Heat Kernel To the HKS

HKS

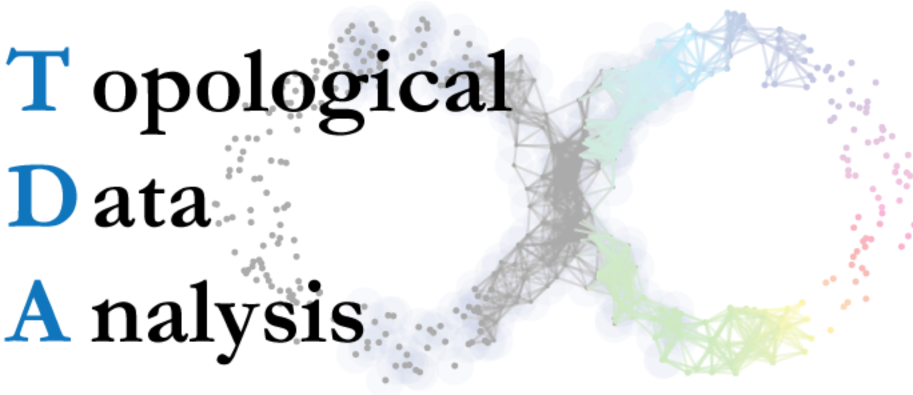
The **Heat Kernel Signature** $K_t(x, x)$ provides a univariate function encoding curvature information (Sun et al, 2009).



(Source: MeshLab)

Topological Data Analysis

We need to capture the relevant curvature information free of location in space.

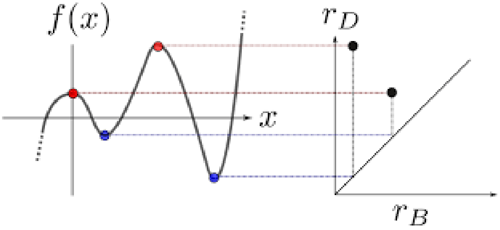


(Source: Topological Data Analysis at CUNY)

(sub)Level Set Filtration

Definition

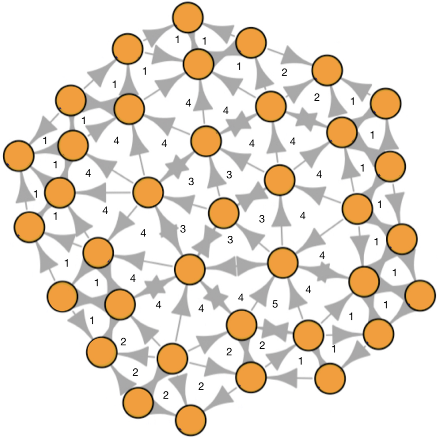
A time t sub level set of real valued function f is defined as the set of $\{x \in f | f(x) \leq t\}$ providing us with a filter for persistence diagrams.



(Source: ResearchGate)

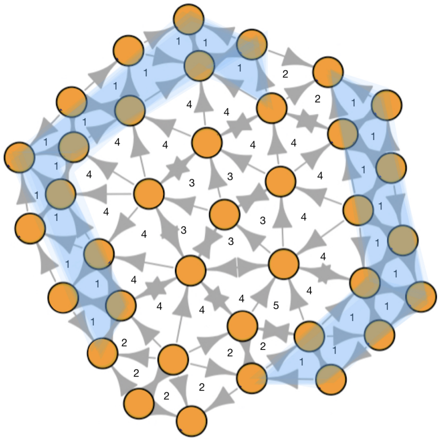
We examine the connected components of the sub level set expanding from local minima.

Sublevel set filtration for function (HKS) defined on simplicial complex



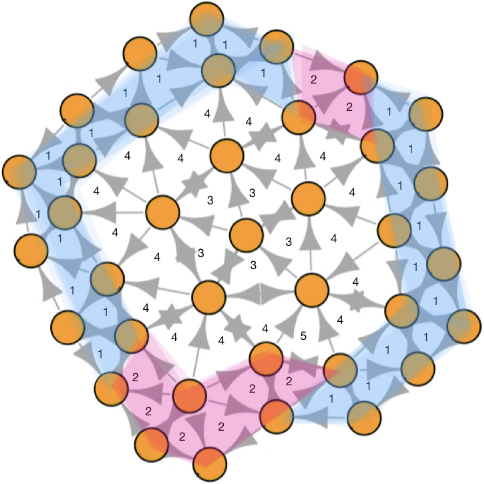
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex



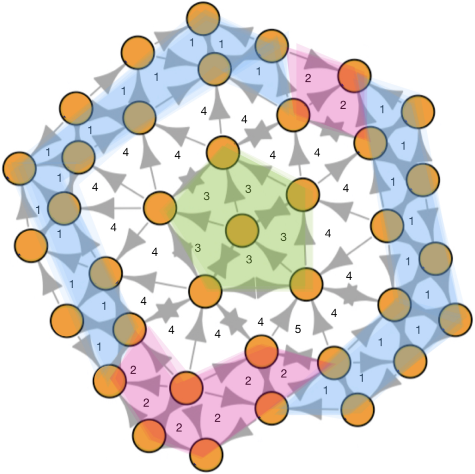
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex



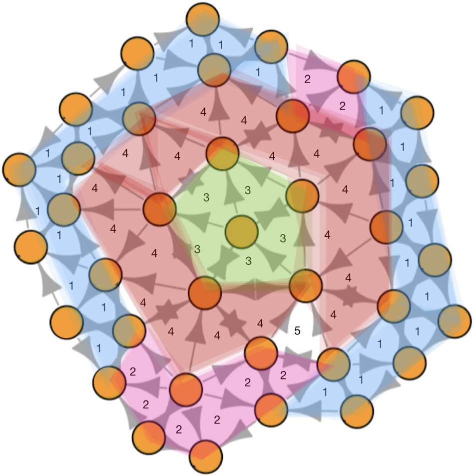
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex



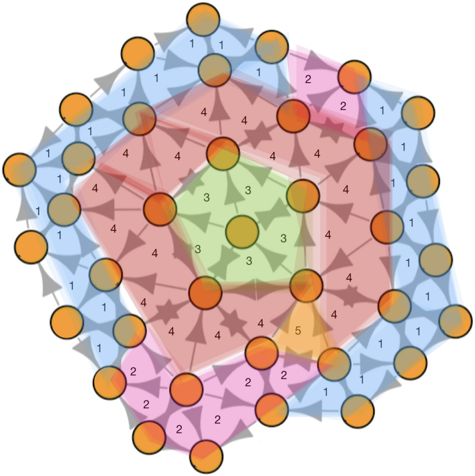
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex



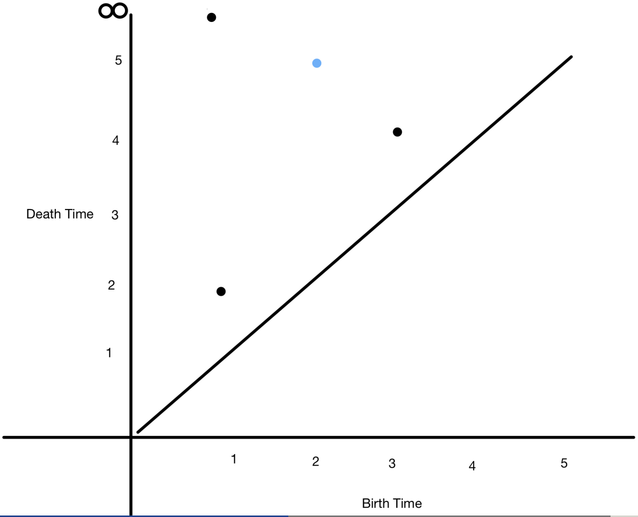
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex



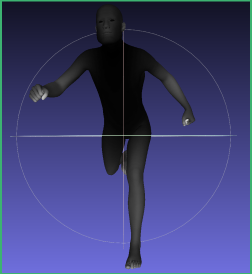
(Source: Rstudio)

Sublevel set filtration for function defined on simplicial complex

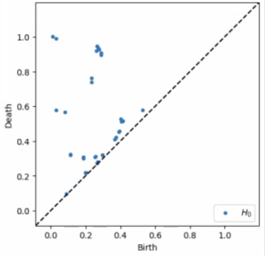


Example of a mesh to persistence diagram

Example of Mesh to Persistence Diagram



→



(Source:MeshLab Rstudio)

Back to Safety

PSSK

These persistence diagrams F are our "data points." One can define a kernel on the space of persistence diagrams $K(F, G) =< \phi(F), \phi(G) >$. We use the Persistence Scale Space Kernel (PSSK) (Reininghaus, et al 2014).

We can use now use SVMs or any of our other standard "kernelizable" learners (kNN, PCA, etc) that we know and love from Eucliden data.

Results

With the synthetic data set, correct classification rates of 99.3 ± 0.9 are achieved.

With real data, 62.7 ± 4.6 were achieved.

This is a 15 class classification problem!

Key Sources

Reininghaus, J., Huber, S., Bauer, U., Kwitt, R. (2015). A stable multi-scale kernel for topological machine learning. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4741-4748.

Sun, J., Ovsjanikov, M., Guibas, L.J. (2009). A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. Computer Graphics Forum, 28.

Thank You!