Topology and Machine Learning
A Global Map of your Data

Anthony Bak

AYASDI
Outline

Part I

- Problems with Big Data
Outline

Part I

- Problems with Big Data
- Topological Summaries

Caveats: I am only talking about the strain of TDA done by Ayasdi
Outline

Part I
- Problems with Big Data
- Topological Summaries
- Examples

Caveats: I am only talking about the strain of TDA done by Ayasdi
Outline

Part I

▶ Problems with Big Data
▶ Topological Summaries
▶ Examples

Part II
Outline

Part I
  ▶ Problems with Big Data
  ▶ Topological Summaries
  ▶ Examples

Part II
  ▶ Review
Outline

Part I

▶ Problems with Big Data
▶ Topological Summaries
▶ Examples

Part II

▶ Review
▶ Why Topology? (Big ideas with examples)
Outline

Part I

▶ Problems with Big Data
▶ Topological Summaries
▶ Examples

Part II

▶ Review
▶ Why Topology? (Big ideas with examples)
▶ More Examples
Outline

Part I

▶ Problems with Big Data
▶ Topological Summaries
▶ Examples

Part II

▶ Review
▶ Why Topology? (Big ideas with examples)
▶ More Examples

Caveats: I am only talking about the strain of TDA done by Ayasdi
Goals

TDA Review
The Data Problem

How do we extract meaning from **Complex Data**?
The Data Problem

How do we extract meaning from **Complex Data**?

- Data is complex because it’s ”Big Data”
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s ”Big Data”
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s ”Big Data”
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
- Or both!
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s "Big Data"
- Or has very rich features (e.g. Genetic Data >500,000 features, complicated interdependencies)
- Or both!
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s ”Big Data”
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
- Or both!

Problem 1: There isn’t a single story happening in your data.
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s ”Big Data”
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
- Or both!

**Problem 1:** There isn’t a single story happening in your data.
**Problem 2:** Too many hypothesis to check.
The Data Problem

How do we extract meaning from **Complex Data**?

- Data is complex because it’s ”Big Data"
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
- Or both!

**Problem 1**: There isn’t a single story happening in your data.
**Problem 2**: Too many hypothesis to check.

TDA will be the tool that summarizes out the irrelevant stories to get at something interesting.
The Data Problem

How do we extract meaning from Complex Data?

- Data is complex because it’s ”Big Data”
- Or has very rich features (eg. Genetic Data >500,000 features, complicated interdependencies)
- Or both!

Problem 1: There isn’t a single story happening in your data.
Problem 2: Too many hypothesis to check.

TDA will be the tool that summarizes out the irrelevant stories to get at something interesting.

The shape (segmentations, groupings) represent verified hypothesis. You have to decide if they are interesting.
Math World

We start in "Math World"
Math World

We start in "Math World"

▶ We’ll draw the data as a smooth manifold.
We start in "Math World"

- We’ll draw the data as a smooth manifold.
- Functions that appear are smooth or continuous.
We start in "Math World"

- We’ll draw the data as a smooth manifold.
- Functions that appear are smooth or continuous.

⇒ We will not need either of these assumptions once we’re in "Data World".
Math World

We start in "Math World"

- We’ll draw the data as a smooth manifold.
- Functions that appear are smooth or continuous.

⇒ We will not need either of these assumptions once we’re in "Data World".

⇒ Even more importantly, data in the real world is never like this.
Math World

Data
Math World

Data

\[ f \]

\[ p \]
\[ f^{-1}(p) \]
\[ f^{-1}(p) \]
Math World

Data

\[ f^{-1}(p) = \Rightarrow q \]

\[ f \]

\[ p \quad \quad \quad \quad q \]
Data

\[ f^{-1}(p) \]

\[ f \]

\[ q \]
Math World

Data

$f^{-1}(p)$

\[ f^{-1}(p) = \Rightarrow q \]

Diagram:

- A diagram showing a function $f$ mapping from $p$ to $q$.
$$f^{-1}(p) \xrightarrow{g} f \rightarrow q \xrightarrow{f} p$$
Math World

\[ f^{-1}(p) = \Rightarrow q \]

Data

\[ f^{-1}(p') \]

\[ g \]

\[ q \]

\[ p \]

\[ \equiv \]
$$f^{-1}(p)$$
Math World

Data

\[ f^{-1}(p) \]

\[ f(p) = q \]

\[ p' \quad q' \]

\[ g \]

\[ f \]

\[ \iff \]

\[ p \quad q \]

Math World

\[ f^{-1}(p) \]

\[ f^{-1}(p) \rightarrow g \]

\[ g \rightarrow f \]

\[ \Rightarrow \]

\[ p' \quad q' \]

\[ p \quad q \]

\[ \Leftrightarrow \]
Math World

\[ f^{-1}(p) \Rightarrow q \]

\[ f^{-1}(p) = \Rightarrow q \]

\[ g \]

Data

\[ r' \]
Math World

Data

\[ f^{-1}(p) = \Rightarrow q \]

\[ g \]

\[ r' \]

\[ f \]

\[ q \]

\[ p \]
Math World

\[ f^{-1}(p) = \Rightarrow q \]

Data

\[ f \]

\[ g \]
Why is this useful?
Why is this useful?

⇒ We get ”easy” understanding of the localizations of quantities of interest.
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?
Why is this useful?

- Lenses inform us where in the space to look for phenomena.
- For easy localizations many different lenses will be informative.
- For hard (= geometrically distributed) localizations we have to be more careful.
- But even then, we frequently get incremental knowledge even from a poorly chosen lens.
Why is this useful?

- Lenses inform us where in the space to look for phenomena.
Why is this useful?

- Lenses inform us where in the space to look for phenomena.
- For easy localizations many different lenses will be informative.
Why is this useful?

- Lenses inform us where in the space to look for phenomena.
- For easy localizations many different lenses will be informative.
- For hard (= geometrically distributed) localizations we have to be more careful.
Why is this useful?

- Lenses inform us where in the space to look for phenomena.
- For easy localizations many different lenses will be informative.
- For hard (= geometrically distributed) localizations we have to be more careful. But even then, we frequently get incremental knowledge even from a poorly chosen lens.
Step 2: Clustering as $\pi_0$

We need to adjust the ”Math World view to bring the algorithm into ”Data World”.
Step 2: Clustering as $\pi_0$

We need to adjust the ”Math World view to bring the algorithm into ”Data World”.

- We replace points with open sets in the range of the lens.
Step 2: Clustering as $\pi_0$

We need to adjust the "Math World view to bring the algorithm into "Data World".

- We replace points with open sets in the range of the lens.
- We replace "connected component of the inverse image" is with "clusters in the inverse image".
Step 2: Clustering as \( \pi_0 \)

We need to adjust the ”Math World view to bring the algorithm into ”Data World”.  

- We replace points with open sets in the range of the lens.
- We replace ”connected component of the inverse image” is with ”clusters in the inverse image”.
- We connect clusters (nodes) with an edge if they share points in common.
Step 2: Clustering as $\pi_0$
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points.

Edges represent shared points between the clusters.

This is also called taking the nerve of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points

Edges represent shared points between the clusters

This is also called taking the nerve of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points
Edges represent shared points between the clusters
This is also called taking the nerve of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points
Edges represent shared points between the clusters

This is also called taking the nerve of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points. Edges represent shared points between the clusters. This is also called taking the nerve of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points

$\Rightarrow$ Nodes are clusters of data points
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens + clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Step 2: Clustering as $\pi_0$

- Nodes are clusters of data points
- Edges represent shared points between the clusters

This is also called taking the *nerve* of a covering where the lens+clustering makes the cover.
Lenses: Where do they come from

The technique rests on finding good lenses.
Lenses: Where do they come from

The technique rests on finding good lenses.

⇒ Luckily lots of people have worked on this problem
Lenses: Where do they come from

A Non Exhaustive Table of Lenses
Lenses: Where do they come from

- Standard data analysis functions

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
</tr>
</thead>
</table>
Lenses: Where do they come from

- Standard data analysis functions

A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions

<table>
<thead>
<tr>
<th>A Non Exhaustive Table of Lenses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistics</strong></td>
</tr>
<tr>
<td>Mean/Max/Min</td>
</tr>
<tr>
<td>Variance</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions

<table>
<thead>
<tr>
<th>A Non Exhaustive Table of Lenses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistics</strong></td>
</tr>
<tr>
<td>Mean/Max/Min</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>n-Moment</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions

<table>
<thead>
<tr>
<th>A Non Exhaustive Table of Lenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
</tr>
<tr>
<td>Mean/Max/Min</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>n-Moment</td>
</tr>
<tr>
<td>Density</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions

---

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>n-Moment</td>
</tr>
<tr>
<td>Density</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology

A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology

---

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
</tr>
<tr>
<td>n-Moment</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology

A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
</tr>
<tr>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A Non Exhaustive Table of Lenses
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

<table>
<thead>
<tr>
<th>A Non Exhaustive Table of Lenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
</tr>
<tr>
<td>Mean/Max/Min</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>n-Moment</td>
</tr>
<tr>
<td>Density</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>Error/Debugging Info</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>Error/Debugging Info</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics
- Domain Knowledge / Data Modeling

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
<th>Data Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
<td>Error/Debugging Info</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics
- Domain Knowledge / Data Modeling

### A Non Exhaustive Table of Lenses

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
<th>Data Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
<td>Age</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
<td></td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>Error/Debugging Info</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics
- Domain Knowledge / Data Modeling

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
<th>Data Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
<td>Age</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
<td>Dates</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>Error/Debugging Info</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Lenses: Where do they come from

- Standard data analysis functions
- Geometry and Topology
- Modern Statistics
- Domain Knowledge / Data Modeling

**A Non Exhaustive Table of Lenses**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Geometry</th>
<th>Machine Learning</th>
<th>Data Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Max/Min</td>
<td>Centrality</td>
<td>PCA/SVD</td>
<td>Age</td>
</tr>
<tr>
<td>Variance</td>
<td>Curvature</td>
<td>Autoencoders</td>
<td>Dates</td>
</tr>
<tr>
<td>n-Moment</td>
<td>Harmonic Cycles</td>
<td>Isomap/MDS/TSNE</td>
<td>...</td>
</tr>
<tr>
<td>Density</td>
<td>...</td>
<td>SVM Distance from Hyperplane</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>Error/Debugging Info</td>
<td>...</td>
</tr>
</tbody>
</table>
Why use TDA?

Basic Example: Higher Fidelity PCA
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

Advantages:
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**
- Provides unsupervised dimensionality reduction.
PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**
- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.
Unsupervised Learning: PCA

PCA captured 98.4% of the variance
Unsupervised Learning: PCA

PCA captured 98.4% of the variance
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**

- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**
- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.

**Problems:**
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**

- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.

**Problems**

- Clusters are distinct but are projected on top of each other.
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**
- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.

**Problems**
- Clusters are distinct but are projected on top of each other.
- Points are falsely clustered because of the projection.
Unsupervised Learning: PCA

PCA is roughly speaking orthogonal projection onto the plane that best contains the data.

**Advantages:**
- Provides unsupervised dimensionality reduction.
- Easy to interpret: Finds the best linear subspace that captures the variance or spread of the data.

**Problems**
- Clusters are distinct but are projected on top of each other.
- Points are falsely clustered because of the projection.

As a *framework* for data analysis we get **higher fidelity** from existing tools.
Real Examples
Supervised Learning: Model Introspection

We can use TDA to examine what is happening with our machine learning models.
Model Introspection: Outliers

**Data:** Customer attributes. Service usages, contractual details.

**Problem:** Customers commit fraud. Find customers with abnormal costs.

**Proposed Solution:** Create an ensemble of cost outlier models. Use these to flag customers as being fraudulent.
Model Introspection: Outliers

TDA Introspection:

- Create a dataset that contains all non-cost information.
- Color by who is being flagged by the ensemble as being a (high) cost outlier.
Model Introspection: Outliers

TDA Introspection:
- Create a dataset that contains all non-cost information.
Model Introspection: Outliers

**TDA Introspection:**
- Create a dataset that contains all non-cost information.
- Color by who is being flagged by the ensemble as being a (high) cost outlier.
Model Introspection: Outliers

Observation:

- The different (independent?) models are all flagging the same group of customers as cost outliers.
Model Introspection: Outliers

Observation:

- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
Model Introspection: Outliers

Observation:

- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
- The client was completely unaware that they were flagging a consistent group of customers. Assumed it was distributed throughout the space.
Model Introspection: Outliers

Observation:
- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
- The client was completely unaware that they were flagging a consistent group of customers. Assumed it was distributed throughout the space.

Further Investigation:
Model Introspection: Outliers

Observation:
- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
- The client was completely unaware that they were flagging a consistent group of customers. Assumed it was distributed throughout the space.

Further Investigation:
- Have we found a model free outlier model? No cost information needed?
Model Introspection: Outliers

Observation:

- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
- The client was completely unaware that they were flagging a consistent group of customers. Assumed it was distributed throughout the space.

Further Investigation:

- Have we found a model free outlier model? No cost information needed?
- More likely: Our models have a systematic bias.
Model Introspection: Outliers

Observation:
- The different (independent?) models are all flagging the same group of customers as cost outliers. Remember that we didn’t use cost in the network creation.
- The client was completely unaware that they were flagging a consistent group of customers. Assumed it was distributed throughout the space.

Further Investigation:
- Have we found a model free outlier model? No cost information needed?
- More likely: Our models have a systematic bias.

⇒ TDA tells us where to look in our data for problems and questions.
Thank You!

http://www.ayasdi.com/
Part II
1. Review

2. Why Topolgy? (With Examples)

3. More Applications
TDA is a machine for creating geometric/topological summaries.
TDA is a machine for creating geometric/topological summaries.

The shape (segmentations, groupings, features) represent verified hypothesis. You have to decide if they are interesting.
Why Topology?

Topology has three properties that make it well suited for data analysis.
Why Topology?

Topology has three properties that make it well suited for data analysis

1. Coordinate Invariance
Why Topology?

Topology has three properties that make it well suited for data analysis

1. Coordinate Invariance
2. Deformation Invariance
Why Topology?

Topology has three properties that make it well suited for data analysis:

1. Coordinate Invariance
2. Deformation Invariance
3. Compressed Representation
Why Topology?

Topology has three properties that make it well suited for data analysis

1. Coordinate Invariance
2. Deformation Invariance
3. Compressed Representation

We’ll examine them in turn.
1) Coordinate Invariance
Coordinate Invariance

- The topology of shape doesn’t depend on the coordinates used to describe the shape.
Coordinate Invariance

- The topology of shape doesn’t depend on the coordinates used to describe the shape.
- Many different feature sets can describe the same phenomena
Coordinate Invariance

- The topology of shape doesn’t depend on the coordinates used to describe the shape.
- Many different feature sets can describe the same phenomena
- While processing data we frequently alter the coordinates: scaling, rotation, whitening
Coordinate Invariance

- The topology of shape doesn’t depend on the coordinates used to describe the shape.
- Many different feature sets can describe the same phenomena
- While processing data we frequently alter the coordinates: scaling, rotation, whitening

⇒ You want to study properties of your data that are invariant under coordinate changes.
Coordinate Invariance: Gene Expression

We want to study a specific biological phenomena via gene expression, such as cancer. We compare the data using:
Coordinate Invariance: Gene Expression

We want to study a specific biological phenomena via gene expression, such as cancer. We compare the data using:

- Samples from different patient populations
Coordinate Invariance: Gene Expression

We want to study a specific biological phenomena via gene expression, such as cancer. We compare the data using:

- Samples from different patient populations
- Different collections of genes
Coordinate Invariance: Gene Expression

We want to study a specific biological phenomena via gene expression, such as cancer. We compare the data using:

- Samples from different patient populations
- Different collections of genes
- Different underlying technology
Coordinate Invariance: Gene Expression

We want to study a specific biological phenomena via gene expression, such as cancer. We compare the data using:

- Samples from different patient populations
- Different collections of genes
- Different underlying technology

⇒ Different coordinates on Cancer
Coordinate Invariance: Gene Expression

NKI

GSE230

ESR1 Levels
2) Deformation Invariance
Deformation Invariance

- Topological features don’t change when you stretch and distort the data
Deformation Invariance

- Topological features don’t change when you stretch and distort the data

**Advantage:** Makes problems easier.
Deformation Invariance

- Topological features don’t change when you stretch and distort the data

**Advantage:** Makes problems easier.

- Noise resistance.
Deformation Invariance

- Topological features don’t change when you stretch and distort the data

**Advantage:** Makes problems easier.

- Noise resistance.
- Less preprocessing of the data.
Deformation Invariance

- Topological features don’t change when you stretch and distort the data

**Advantage**: Makes problems easier.

- Noise resistance.
- Less preprocessing of the data.
- Robust (stable) answers.
Deformation Invariance: Line and Noisy Line

Pearson Correlation: 0.999998 resp 0.9999

Use x-axis coordinate as a lens. Expect that we will get two lines out.
Deformation Invariance: Line and Noisy Line

Pearson Correlation: 0.999998 resp 0.9999

Use x-axis coordinate as a lens. Expect that we will get two lines out.
Deformation Invariance: Line is a Circle
Some lessons.

- We can be surprised even when we think the solution is obvious. Both examples had almost perfect correlation.
Deformation Invariance: Circle and Noisy Line

Some lessons.

- We can be surprised even when we think the solution is obvious. Both examples had almost perfect correlation.
- We did not think to transform the data to look for structure. TDA saved us.
Deformation Invariance: Circle and Noisy Line

Some lessons.

- We can be surprised even when we think the solution is obvious. Both examples had almost perfect correlation.
- We did not think to transform the data to look for structure. TDA saved us.
- Being insensitive to deformation means we discover unexpected structure.
Deformation Invariance: Circle and Noisy Line

Some lessons.

- We can be surprised even when we think the solution is obvious. Both examples had almost perfect correlation.
- We did not think to transform the data to look for structure. TDA saved us.
- Being insensitive to deformation means we discover unexpected structure.
- We did not find structure in noise.
Deformation Invariance: Intertwined Spirals

Separate the two classes.
Deformation Invariance: Intertwined Spirals

Separate the two classes.

Use x-axis coordinate as a lens.
Deformation Invariance: Intertwined Spirals
Deformation Invariance: Intertwined Spirals

Some lessons.
- Separating the two classes was easy. Take connected components of graph.
Deformation Invariance: Intertwined Spirals

Some lessons.

- Separating the two classes was easy. Take connected components of graph.
- We retained more information than clustering.
Deformation Invariance: Intertwined Spirals

Some lessons.
- Separating the two classes was easy. Take connected components of graph.
- We retained more information than clustering. We remember that we have lines.
Deformation Invariance: Intertwined Spirals

Some lessons.

- Separating the two classes was easy. Take connected components of graph.
- We retained more information than clustering. We remember that we have lines.
- If there was localized structure along the spiral, for example, subclasses of the two major classes, we would find those localizations on these lines.
3) Compressed Representation
Compressed Representation

- Replace the metric space with a combinatorial summary: a simplicial complex.
Compressed Representation

- Replace the metric space with a combinatorial summary: a simplicial complex.
- The data is easier to manage, search and query while maintaining essential features.
Compressed Representation

- Replace the metric space with a combinatorial summary: a simplicial complex.
- The data is easier to manage, search and query while maintaining essential features.
- Leverage many known algorithms from Graph Theory, Computational Topology and Computational Geometry
Compressed Representation

- Replace the metric space with a combinatorial summary: a simplicial complex.
- The data is easier to manage, search and query while maintaining essential features.
- Leverage many known algorithms from Graph Theory, Computational Topology and Computational Geometry

This is more or less what TDA is about
Analogy: Cartography

To be a good and useful map:

▶ Come at your problem with as few assumptions as possible but bring tools to measure what’s there (metrics & lenses)
Analogy: Cartography

To be a good and useful map:

- Come at your problem with as few assumptions as possible but bring tools to measure what’s there (metrics & lenses)
- Measure what you find. Use as few assumptions as possible.
Analogy: Cartography

To be a good and useful map:

- Come at your problem with as few assumptions as possible but bring tools to measure what’s there (metrics & lenses).
- Measure what you find. Use as few assumptions as possible.
- Produce a summary relevant to the problem.
Analogy: Cartography

To be a good and useful map:

- Come at your problem with as few assumptions as possible but bring tools to measure what’s there (metrics & lenses).
- Measure what you find. Use as few assumptions as possible.
- Produce a summary relevant to the problem.
  - Different problems require different summaries.
Analogy: Cartography

To be a good and useful map:

▶ Come at your problem with as few assumptions as possible but bring tools to measure what’s there (metrics & lenses)
▶ Measure what you find. Use as few assumptions as possible.
▶ Produce a summary relevant to the problem.
   ▶ Different problems require different summaries.

Use your map to make decisions! Don’t got back and measure from scratch.

⇒ TDA is the machine that takes the tools (metrics & lenses) and produces the summary (network)
More Examples
Customer Churn

**Data**: Customer usage and contractual details for major telco.

**Analysis**: A contractual stage data lens was used to split the data into "contractual stage" groups.
Customer Churn

Shape and Meaning
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
- We can see natural segmentations that vary over many orders of magnitude (100-50,000 customers) in size.
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
- We can see natural segmentations the vary over many orders of magnitude (100-50,000 customers) in size.
- Stability gives us confidence in the validity of the results.
Customer Churn

Shape and Meaning

▶ We see similar shape across all the contract stages.
▶ We can see natural segmentations the vary over many orders of magnitude (100-50,000 customers) in size.
▶ Stability gives us confidence in the validity of the results.
▶ The shape tells us where to look in the data. We were able to localize churn in certain contract stages.

Coloring helps us figure out what is going on.

⇒ We turn our insight into better targeting resulting in fewer lost customers.

This can be automated.
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
- We can see natural segmentations the vary over many orders of magnitude (100-50,000 customers) in size.
- Stability gives us confidence in the validity of the results.
- The shape tells us where to look in the data. We were able to localize churn in certain contract stages.
- Coloring helps us figure out what is going on.
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
- We can see natural segmentations that vary over many orders of magnitude (100-50,000 customers) in size.
- Stability gives us confidence in the validity of the results.
- The shape tells us where to look in the data. We were able to localize churn in certain contract stages.
- Coloring helps us figure out what is going on.
- Not shown: The software also allows for statistical queries of subgroups of data points.

⇒ We turn our insight into better targeting resulting in fewer lost customers. This can be automated.
Customer Churn

Shape and Meaning

- We see similar shape across all the contract stages.
- We can see natural segmentations the vary over many orders of magnitude (100-50,000 customers) in size.
- Stability gives us confidence in the validity of the results.
- The shape tells us where to look in the data. We were able to localize churn in certain contract stages.
- Coloring helps us figure out what is going on.
- Not shown: The software also allows for statistical queries of subgroups of data points.

⇒ We turn our insight into better targeting resulting in fewer lost customers.
Customer Churn

Shape and Meaning

► We see similar shape across all the contract stages.
► We can see natural segmentations that vary over many orders of magnitude (100-50,000 customers) in size.
► Stability gives us confidence in the validity of the results.
► The shape tells us where to look in the data. We were able to localize churn in certain contract stages.
► Coloring helps us figure out what is going on.
► Not shown: The software also allows for statistical queries of subgroups of data points.

⇒ We turn our insight into better targeting resulting in fewer lost customers. This can be automated.
Predictive Maintenance: Industrial Machinery

**Setup**: We have a large piece of industrial machinery, think turbine, jet engine, locomotive or robot. Built into the machine are sensors measuring physical quantities: pressure, temperature, rpms etc.

**Problem**: Unscheduled downtime is very expensive.

**Question**: Can we predict when a part will need to be repaired in the future so we can schedule the downtime appropriately?
Data Transformation: We want the sensors to be comparable. In this example, z-scoring is sensible.
**Data Transformation**: We want the sensors to be comparable. In this example, z-scoring is sensible (but there are other sensible choices as well, min/max normalization, logs if sensors vary of several orders of magnitude).
Predictive Maintenance: Industrial Machinery

High mean, high variance

High mean, low variance
Fraud Detection

About the data:

600,000+ transactions for a given month
Each transaction has 140 attributes (account, device, timing)
Fraudulent transactions that were not caught were flagged by chargebacks
Fraud Detection

Wherever the network lights up is a failure of the rules engine.
Fraud Detection

Occurrence of Credit Charge Back

<table>
<thead>
<tr>
<th>Category Code</th>
<th>Occurrence Count</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1234</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>5678</td>
<td>High</td>
</tr>
</tbody>
</table>

Enriched for the following attributes:
1. Images disabled
2. Javascript disabled
3. Cookies disabled
4. Flash disabled
5. Time spent on page was significantly longer
Emergency room triage model

Predicted mortality
Low  |  High

Actual mortality
Low  |  High

Missing responses to particular questions caused model to fail for these patients
Parkinson’s Detection with Mobile Phone
Oil Well Sensors and Recovery

Well Production
Low
High

Low output wells

High output wells

High output wells in a moderately producing region
Analyzing NGS Data with Ayasdi Cure

**About the Data**
- 164 patients from autism clinical trial
- Some with autism, some without
- Data consists of genotype calls

**Goal:** Identify genetic drivers of the disease in subpopulations
Analyzing NGS Data with Ayasdi Cure

Patients in the trial with Autism

Disease Phenotype for Autism

High

Low
Analyzing NGS Data with Ayasdi Cure
Variants with association to Crohn's Disease

Low

High
Isomap: Configuration Space of $\text{C}_8\text{H}_{16}$
Malware System Calls
User Experience for Care Paths

Drag and Drop Interface

Care Path Overview

Patient Query

Patient Query Detail
What’s the point of all this?
Data Has Shape
And Shape Has Meaning
Thank You!

http://www.ayasdi.com/