Mar 23

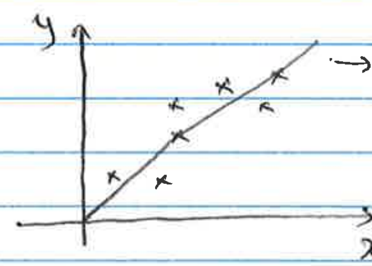
Mapper : Useful in multi-variate data analysis

$f_1 : \mathbb{X} \to \mathbb{R}$: f_1 is scalar valued function
$f_2 : \mathbb{X} \to \mathbb{R}^d$: f_2 is vector valued function.
Alternatively, we can construct a vector of multiple scalar functions $f = (f_1, f_2, \dots, f_d)^T$

If $\mathbb{X}$ is a point cloud or subset of $\mathbb{R}^d$ $(d > 1)$ then the domain is High Dimensional.
→ We'll make the distinction where high-dimensional ⇒ domain
multi-variate ⇒ range for a function.

\# Let $\mathbb{X}$: high-dim. point cloud with some metric / distance d
$x, y \in \mathbb{X}$, $d(x,y)$ is computed. For this data set we can perform
⇒ Clustering : K-means, K-median, hierarchical / graph-based
⇒ Dimensionality Reduction : PCA, Factor analysis, eigenmap, LLE JLT etc.

When we have $y = f(x)$ $f : \mathbb{X} \to \mathbb{R}^d$ $d \geq 1$
we can perform regression
⇒ linear regression
⇒ logit
⇒ multivariate regression



→ best fitting line to given samples $f : \mathbb{R} \to \mathbb{R}$

\# Mapper : discretization of Reeb graph / Reeb space, "Clustering"
→ easily integrated into machine learning.

X : topological space (point cloud, manifold)

$\mathcal{U}$: open cover of X. Collection of open sets

$$\mathcal{U} = \{U_\alpha\}_{\alpha \in A} \rightarrow \text{finite index set}$$

s.t.

$$\bigcup_{\alpha \in A} U_\alpha = X \quad (\text{assume } U_\alpha \text{ is path-connected})$$

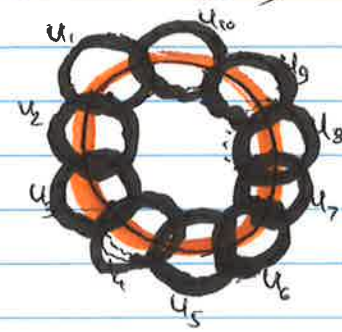
(union of U_α "covers" X)

example:

$X = S^1 \subseteq \mathbb{R}^2$ (cycle)

$A = \{1, 2, \ldots, 10\}$

$$\bigcup_{\alpha \in A} (U_\alpha \cap X) = X$$

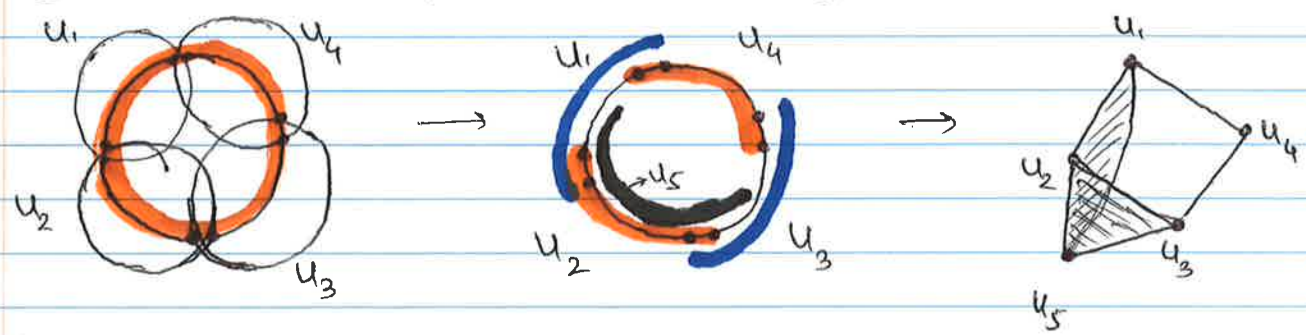


$\rightarrow$ Alternatively, $X \subseteq \bigcup_{\alpha \in A} U_\alpha$

Def: Nerve of a cover : Given a finite cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of X, the nerve of cover $\mathcal{U}$ is the simplicial complex $N(\mathcal{U})$ whose vertex set is the index set A and a subset $\{\alpha_0, \alpha_1, \ldots, \alpha_k\}$ of A spans a k-simplex in $N(\mathcal{U})$ iff

$$U_{\alpha_0} \cap U_{\alpha_1} \cap \cdots \cap U_{\alpha_k} \neq \emptyset$$

(intersection of open sets indexed by $\alpha_0, \ldots, \alpha_k$ is not empty)



Čech Complex : Nerve complex of a collection of open <u>Balls</u>

Def: [Mapper] Let X, Z be topological spaces and $f: X \to Z$ be a well-behaved continuous function ($Z = \mathbb{R}$ or $\mathbb{R}^2$ for example.)
Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be a finite open cover of Z

The mapper (mapper construction) is defined to be the nerve complex of the pull-back cover $M(\mathcal{U}, f) = N(f^*(\mathcal{U}))$

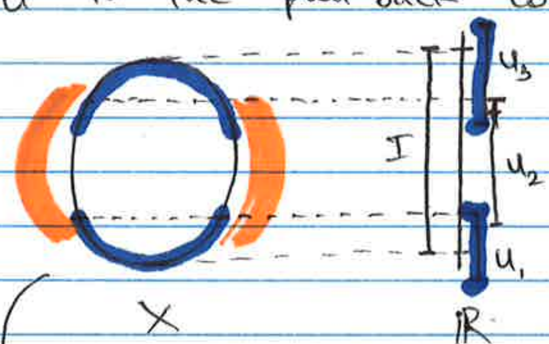
Def: pull-back cover. Let $f: X \to Z$ as above. and Z be equipped with cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$.

The sets $\{f^{-1}(U_\alpha)\}_{\alpha \in A}$ form an open cover of X

Consider the path connected components of $f^{-1}(U_\alpha)$:

$$f^{-1}(U_\alpha) = \bigcup_{i=1}^{j_\alpha} V_{\alpha,i} \qquad j_\alpha = \# \text{ path-connected components of } f^{-1}(U_\alpha)$$

$f^*(\mathcal{U})$ is the cover of X obtained from cover $\mathcal{U}$ of Z
it is the pull back cover of X (induced by $\mathcal{U}$ via f^{-1})

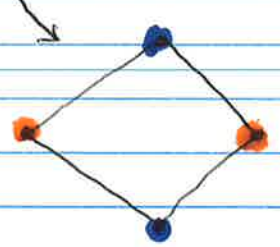


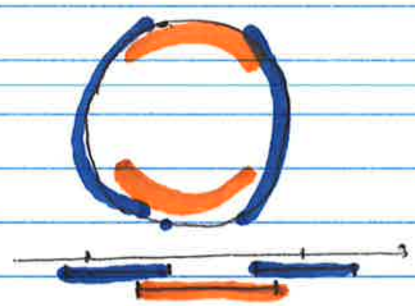
$\mathcal{U} = \{U_1, U_2, U_3\}$

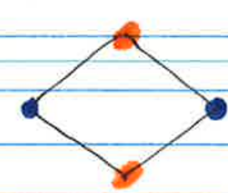
$\bigcup_{i=1,2,3} U_i \supseteq I \subseteq \mathbb{R}$

Suppose $U_1 = (0, 3)$ $\quad I = (1, 10)$
$U_2 = (2, 7)$
$U_3 = (6, 10)$

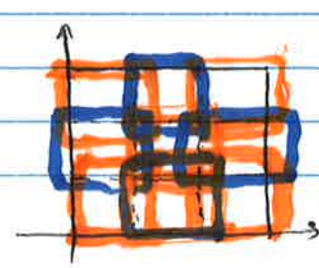
mapper

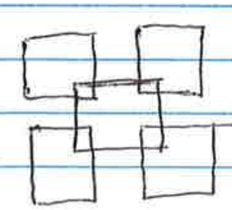


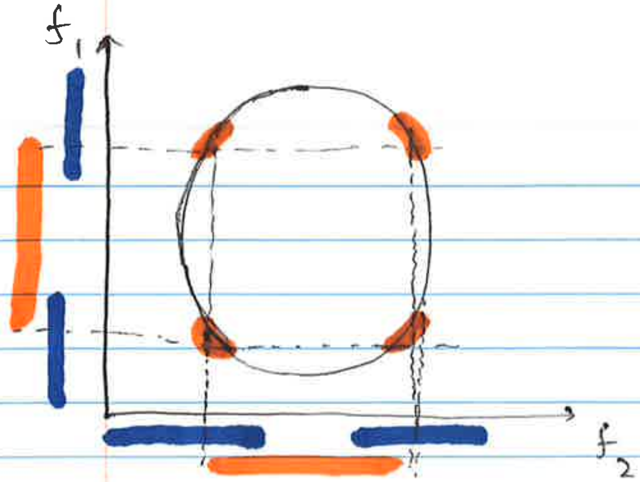


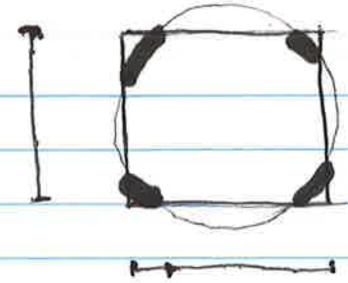


for $f: X \to Z$, $Z \subseteq \mathbb{R}^2$









→ intersection of inverse images of f_1 and f_2 for the specified interval along each dimension.

<u>Def</u>: Well-behaved function f: for every path-connected open set $U \subseteq Z$, $f^{-1}(U)$ has finitely many path connected components

eg. piecewise linear real valued function on a finite S.C.

Mapper gives a "soft clustering" of domain (clusters overlap) but it also gives a relationship between clusters. and it can preserves topological structure (depending on the cover we define)