Suppose X is an unknown metric space. Suppose given a finite sample $S \subset X$ and the distance

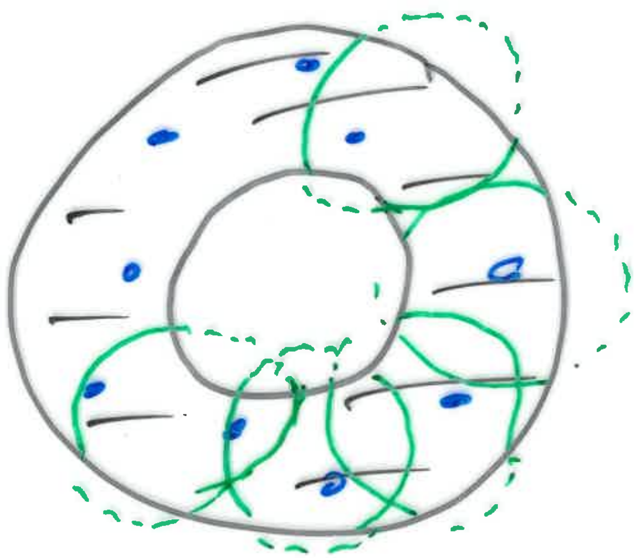
$$d_X(x, y)$$

for all $x, y \in S$.

<u>Problem</u> we'd like to infer something about the topology of X using the distances $d_X(x,y)$ for $x, y \in S$.

<u>Example</u> A sample S is taken from an unknown metric space $X \subseteq \mathbb{R}^2$

X :



S finite sample

If we could construct the ball

$$B(s, \varepsilon) = \{ x \in X : d_X(x, s) < \varepsilon \}$$

we could hope that the collection

$$\mathcal{U}_\varepsilon = \{ B(s, \varepsilon) \}_{s \in X}$$

might be an open cover of X for suitable $\varepsilon > 0$.

If $\mathcal{U}_\varepsilon$ is an open cover of X then Leray's theorem says the nerve $N\mathcal{U}_\varepsilon$ is homotopy equivalent to X (so has the same Euler characteristic etc. as X).

However, we can't construct the balls $B(s, \varepsilon)$ as we don't know X.

Recall: $N\mathcal{U}_\varepsilon$ is a simplicial complex with one vertex s for each ball $B(s, \varepsilon)$, $s \in S$.

$N\mathcal{U}_\varepsilon$ has an edge

if $B(s, \varepsilon) \cap B(t, \varepsilon) \neq \emptyset$. This will be non-empty if and only if $d_X(s,t) < 2\varepsilon$.

So we can construct the vertices and edges of $N\mathcal{U}_\varepsilon$.

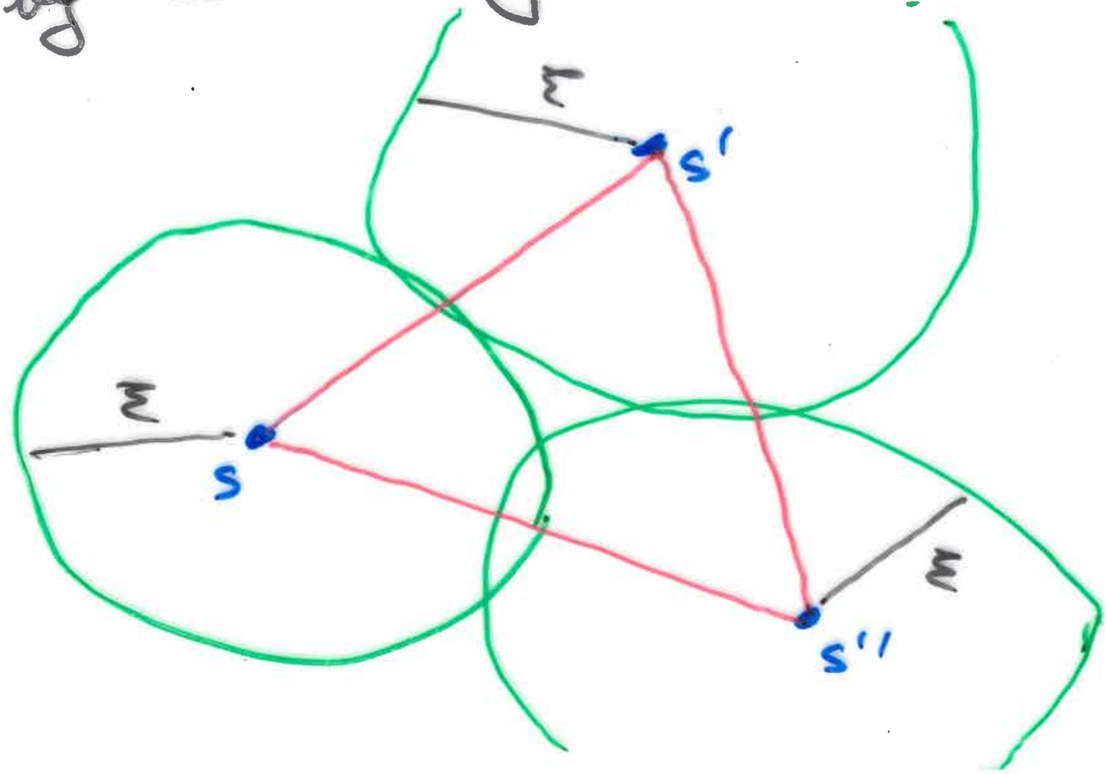
$N\mathcal{U}_\varepsilon$ has a 2-simplex for

$$B(s, \varepsilon) \cap B(s', \varepsilon) \cap B(s'', \varepsilon) \neq \emptyset$$

$s, s', s'' \in S$.

If we happen to know that $X \subseteq \mathbb{R}^n$ then we could compute these balls and their intersections. However, this might be a lengthy computation.

we could approximate $N\mathcal{U}_\varepsilon$ by considering the picture:



We could approximate the nerve $N\mathcal{U}_\varepsilon$ by the clique complex $K_{2\varepsilon}$:

$K_{2\varepsilon}$ has a k-simplex

$$\sigma = \{s_0, s_1, \dots, s_k\}$$

whenever

$d(s_i, s_j) < 2\varepsilon$ for all $s_i, s_j \in \sigma$.

<u>Proposition</u>

$$N\mathcal{U}_\varepsilon \subseteq K_{2\varepsilon} \subseteq N\mathcal{U}_{2\varepsilon}$$

Conclusion:

K_ε is a good approximation to $N\mathcal{U}_\varepsilon$ since we consider a range of $\varepsilon > 0$.

Illustration: Natural Images

van Hateren & van der Schaaf: 4167 digital photos of random outdoor scenes.

Illustration: Natural Images

van Hateren & van der Schaaf: 4167 digital photos of random outdoor scenes.

Mumford, Lee, Pedersen: 8 000 000 random high-contrast 3×3 patches in $\mathbb{E}^9$.

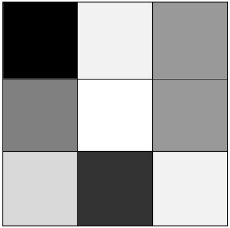
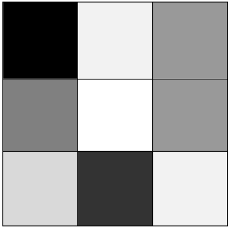


Illustration: Natural Images

van Hateren & van der Schaaf: 4167 digital photos of random outdoor scenes.

Mumford, Lee, Pedersen: 8 000 000 random high-contrast 3×3 patches in $\mathbb{E}^9$.



By normalizing with respect to brightness and contrast the patches are projected onto a set $\mathcal{S}$ of points in a topological seven-sphere $\mathbb{S}^7 \subset \mathbb{E}^8$.

Carlsson, Ishkanov, de Silva, Zomorodian: consider high-density subsets $\mathcal{S}(k) \subset \mathcal{S}$.

Carlsson, Ishkanov, de Silva, Zomorodian: consider high-density subsets $\mathcal{S}(k) \subset \mathcal{S}$.

Codensity function: $\delta_k(x)$ is the Euclidean distance from $x \in \mathcal{S}$ to its kth nearest neighbour in $\mathcal{S}$.

Carlsson, Ishkanov, de Silva, Zomorodian: consider high-density subsets $\mathcal{S}(k) \subset \mathcal{S}$.

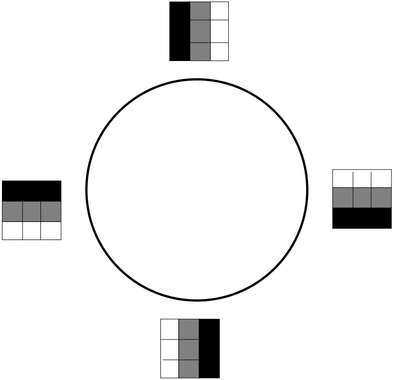
Codensity function: $\delta_k(x)$ is the Euclidean distance from $x \in \mathcal{S}$ to its kth nearest neighbour in $\mathcal{S}$.

$\mathcal{S}(k)$ contains the 25% of patches with least $\delta_k(x)$.

The simplicial complex K_ϵ for $\mathcal{S}(300)$ 'look like' a single circle.

The simplicial complex K_ϵ for $\mathcal{S}(300)$ 'look like' a single circle.
(Later we'll be more precise: The β_1 bar code for a random sampling of $\mathcal{S}(300)$ yields a single persistent homology class.)

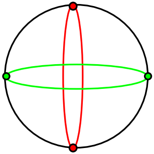
The simplicial complex K_ϵ for $\mathcal{S}(300)$ 'look like' a single circle.
(Later we'll be more precise: The β_1 bar code for a random sampling of $\mathcal{S}(300)$ yields a single persistent homology class.)

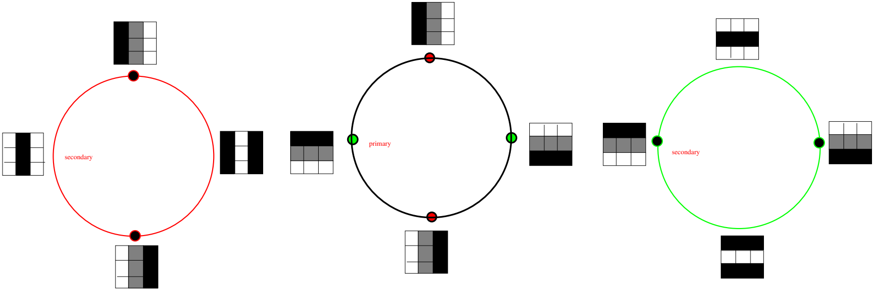


The simplicial complex K_ϵ for $\mathcal{S}(300)$ 'comprises' five circles.

The simplicial complex K_ϵ for $\mathcal{S}(300)$ ‘comprises’ five circles.
(Later we’ll be more precise: The β_1 bar code for a random sampling of $\mathcal{S}(15)$ yields five persistent homology classes.)

The simplicial complex K_ϵ for $\mathcal{S}(300)$ ‘comprises’ five circles.
(Later we’ll be more precise: The β_1 bar code for a random sampling of $\mathcal{S}(15)$ yields five persistent homology classes.)

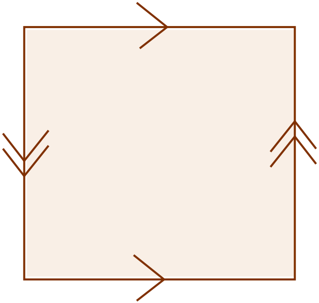




Further inspection suggests that K_ϵ 'looks like' a Klein bottle.
(Later we'll be more precise: The β_2 bar codes are less robust, but seem to indicate a single homology class with mod 2 coefficients.)

Further inspection suggests that K_ϵ 'looks like' a Klein bottle.
(Later we'll be more precise: The β_2 bar codes are less robust, but seem to indicate a single homology class with mod 2 coefficients.)

$$H_2(\text{Klein bottle}, \mathbb{Z}_2) = \mathbb{Z}_2$$



Further inspection suggests that K_ϵ 'looks like' a Klein bottle.
(Later we'll be more precise: The β_2 bar codes are less robust, but seem to indicate a single homology class with mod 2 coefficients.)

$$H_2(\text{Klein bottle}, \mathbb{Z}_2) = \mathbb{Z}_2$$