Classification and Topology

(or Consequences of Sobolev Consistency)

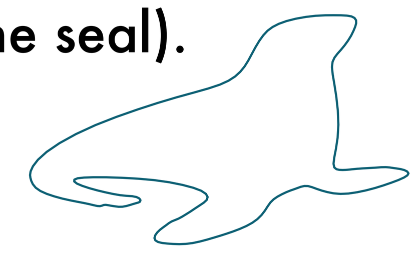
Olivier Binette
Université du Québec à Montréal

Key points

- Topologically accurate class boundary reconstruction in binary classification.
- A fruitful relationship with topology and real algebraic geometry is highlighted.
- Hints at a research project: Can the technique be adapted to different surface reconstruction frameworks? Can similar results be obtained in Bayesian nonparametrics or with neural nets?

The Problem

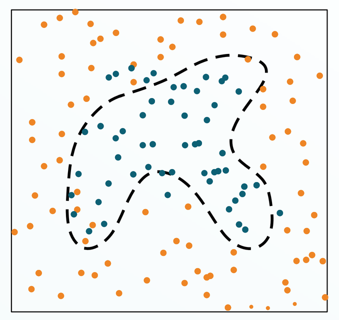
Starting point: $\gamma \subset (0,1)^k$ a compact connected hypersurface (think of the seal).



By the Jordan-Brouwer separation theorem, *γ separates the space in two connected components, one of which is bounded. These components are called* classes *and γ is the* separation surface *(or class boundary).*

A priori, γ is unknown and we are trying to reconstruct it using the following labelled points indexed by $i \in \{1, 2, \dots, n\}$:

$$\ell_i \mid x_i \sim \mathrm{Ber}(p(x_i))$$
$$x_i \text{ i.i.d. uniform}((0,1)^k)$$



where

A1: $p : (0,1)^k \to [0,1]$ is smooth and has regular value 1/2;

A2: $p > 1/2$ in the interior of γ and $p < 1/2$ in the exterior of γ.

Note: $\gamma = p^{-1}(1/2)$

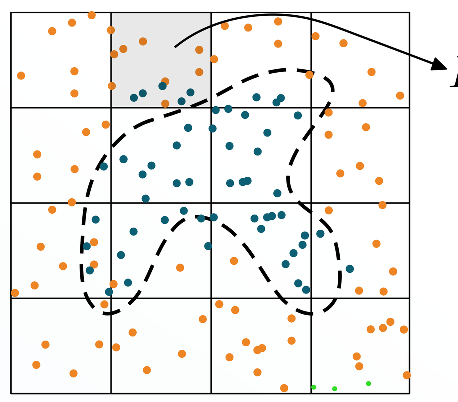
A Solution

By the Nash-Tognoli theorem of real algebraic geometry, compact surfaces are not topologically distinguishable from compact non-singular real algebraic varieties. This suggest estimating γ by $f^{-1}(1/2)$ where $f : (0,1)^k \to \mathbb{R}$ is a polynomial function.

This is done by first estimating p globally using a two-step method.

1. Histogram approximation over the regions

$$R^d_{i,j} = [\tfrac{i}{d+1}, \tfrac{i+1}{d+1}] \times [\tfrac{j}{d+1}, \tfrac{j+1}{d+1}]$$



$$E^4_{1,2} = \frac{5+1}{13+2} \approx (4+1)^2 \int_{R^4_{i,j}} p(\mathbf{x})d\mathbf{x}$$

2. Bernstein polynomial smoothing:

$$\hat{f}_d(u,v) = (d+1)^2 \sum_{i,j=1}^{d} E^d_{i,j} B_{i,d}(u) B_{j,d}(v)$$

where

$$B_{i,d}(u) = \binom{d}{i} u^i (1-u)^{d-i}.$$

We let the degree $d \to \infty$ as the sample size increase, and our estimate of the surface γ is

$$\hat{\gamma} = \hat{f}_d^{-1}(1/2).$$

But does it converge? In what sense?

Main results

Let

$$W^{2,\infty} = \{f : [0,1]^k \to \mathbb{R} \mid \|f\|_{W^{2,\infty}} < \infty\}$$

be a Sobolev space of twice continuously differentiable functions endowed with its norm

$$\|f\|_{W^{2,\infty}} = \|f\|_\infty + \|\nabla f\|_\infty.$$

Theorem

Let $p \in W^{2,\infty}$ have regular value 0 with $p^{-1}(0) \subset (0,1)^k$ and let $\hat{f}_d$ be a $W^{2,\infty}$-consistent* estimator of p. Then a.s.:

1. $\hat{f}_d^{-1}(0)$ converges to $p^{-1}(0)$ in the Hausdorff distance as $d \to \infty$,
2. $\hat{f}_d^{-1}(0)$ is eventually diffeomorphic to $p^{-1}(0)$

Hausdorff distance is:

$$d_H(A,B) = \max\left\{\sup_{x\in A} d(x,B), \sup_{y \in B} d(A,y)\right\}.$$

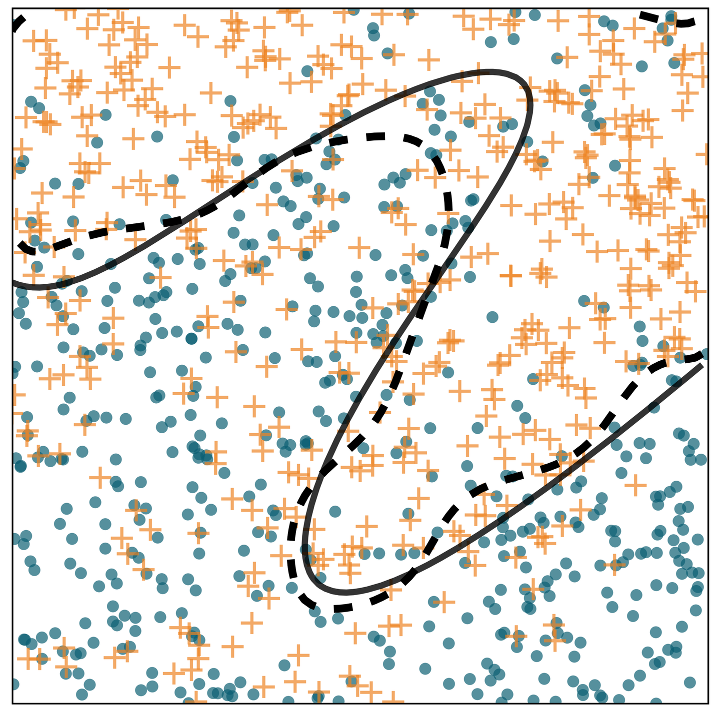
Corollary

If the degree d grows sufficiently slowy in the number of observations, then our estimator $\hat{\gamma}$ is consistent in the Hausdorff + diffeomorphic topology of compact manifold space.

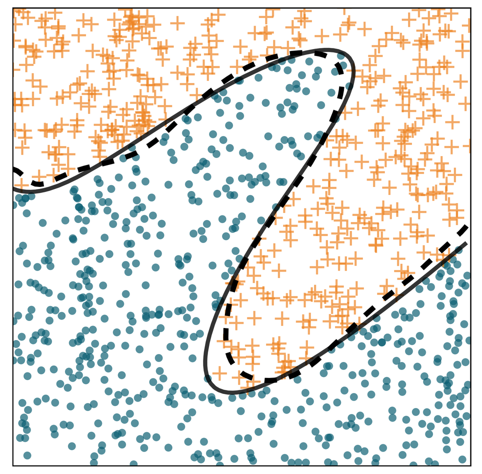
* i.e. $\hat{f}_d$ almost surely converges to p in the Sobolev norm.

Examples

Binary classification:

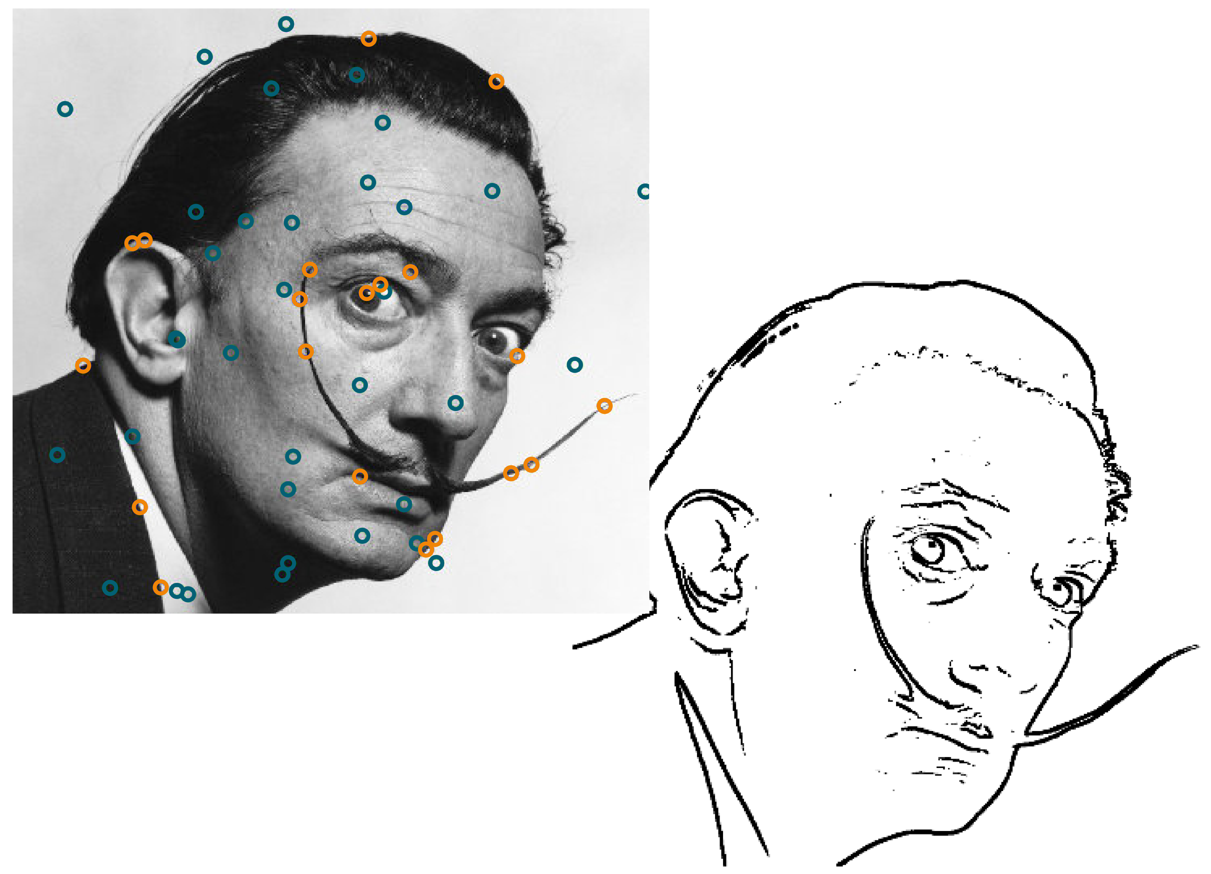


noise



no noise

Supervised edge detection:



References

- Binette, O. (2018) Topologie et apprentissage machine. Dans le *Notes from the Margin*.
- Lima, E.K. (1988) The Jordan-Brouwer separation theorem for smooth hypersurfaces. *Amer. Math. Monthly* 95 (1)
- Telyakovskii, S. A. (2008) On the approximation of differentiable functions by Bernstein polynomials and Kantorovich polynomials. *Proc. Steklov Inst. Math.* 260 (1)
- Kollar, J. (2017) Nash's Work in Real Algebraic Geometry. *Bull. Amer. Math. Soc.* (N.S.) 54 (2)