

Clustering of Distributions on 1-Dimensional Manifolds

Setup

Suppose we have two manifolds \mathcal{M}_1 and $\mathcal{M}_2 \in \mathbb{R}^d$, each of length 1, defined by $f_1(t)$ and $f_2(t)$ respectively ($f_i : [0, 1] \mapsto \mathbb{R}^d$). Define δ as the minimum distance between the two manifolds, i.e., $\delta = \min_t \|f_1(t) - f_2(t)\|$, and let $\delta > 0$. For now, restrict each f_i such that the distance along the manifold between $f_i(t)$ and $f_i(s)$ is equal to the difference between t and s (this also implies that each manifold is of length 1). We sample $T_1, \dots, T_n \stackrel{iid}{\sim} F$ for continuous F with support $[0, 1]$ and use f_1 to map the first n_1 points to \mathcal{M}_1 and f_2 to map the remaining $n_2 = n - n_1$ points to \mathcal{M}_2 . Let $X_i = f_1(T_i)$ and $Y_j = f_2(T_j)$. Without loss of generality, assume $n_1 \leq n_2$.

Theory

Let $D_i = X_{(i)} - X_{(i-1)}$. Then if $\max_i D_i < \delta$, we have sufficient separation of points in \mathcal{M}_1 .

Uniform case

It can be shown that if $X_i \stackrel{iid}{\sim} \text{Uniform}(0, 1)$, then $D_i \sim \text{Beta}(1, n)$. Therefore, $P(\max_i D_i < \delta) \geq (P(D_i < \delta))^n = (1 - (1 - \delta)^n)^n$, which is a decreasing function for sufficiently large n . This gives us the result $\max_i D_i \xrightarrow{a.s.} 0$.

General case

Proof sketch that $P(D_i \leq \delta) \rightarrow 1$

If F is absolutely continuous, then it can be shown that

$$P(D_i \leq \delta) = 1 - \int_0^{1-\delta} \frac{n!}{(n-i+1)!(i-2)!} (F(x))^{i-2} (1 - F(x + \delta))^{n-i} f(x) dx$$

Making the approximation $F(x + \delta) \approx F(x) + \delta f(x)$ and bounding $f(x) \geq a > 0$, we get:

$$P(D_i \leq \delta) \geq 1 - \int_0^{1-\delta} \frac{n!}{(n-i+1)!(i-2)!} (F(x))^{i-2} (1 - F(x) - a\delta)^{n-i} f(x) dx$$

Then making the substitution $u = F(x) \implies du = f(x)dx$, we get

$$1 - \int_0^{F(1-\delta)} \frac{n!}{(n-i+1)!(i-2)!} u^{i-2} (1 - u - a\delta)^{n-i} du$$

Which becomes $1 - CP(n)\delta^n$ for some $C > 0$ and $P(n)$ is a polynomial of n , giving us the desired result $D_i \xrightarrow{P} 0$.

Computational Results

TBD