An introduction to Topological Data Analysis with Gudhi
TP 3:
MVA 2017-18

Frédéric Chazal

December 11, 2017

Abstract

To download this file and get the code samples (available on the course webpage soon):
`http://geometrica.saclay.inria.fr/team/Fred.Chazal/MVA2017.html`

Documentation for the Python interface of Gudhi:
`http://gudhi.gforge.inria.fr/python/latest/`

1 Computing approximation of persistence landscapes

Exercise 1. Write a function to compute the landscape of a persistence diagram discretized on a regular grid.

```
def landscapes_approx(diag,p_dim,x_min,x_max,nb_nodes,nb_ld):
    """Compute a dicretization of the first nb_ld landscape of a
    p_dim-dimensional persistence diagram on a regular grid on the
    interval [x_min,x_max]. The output is a nb_ld x nb_nodes numpy
    array
    + diag: a persistence diagram (in the Gudhi format)
    + p_dim: the dimension in homology to consider
    """

    landscape = np.zeros((nb_ld,nb_nodes))

    ####
    ####  Your code here
    ####

    return landscape
```

Test your function on a few examples of diagrams.

```
#To plot the landscape
L =  landscapes_approx(diag,p_dim,x_min,x_max,nb_nodes,nb_ld)
plt.plot(np.linspace(x_min,x_max, num=nb_nodes),L[0:nb_ld,:].transpose())
```

Exercise 2. Compute and store the persistence landscapes of the accelerometers times series data from TP2. Use the obtained landscape to experiment supervised and non supervised classification on this data.

2 Bootstrap and confidence bands for lanscapes

Exercise 3. The goal of this exercise is to implement the bootstrap algorithm below from [F. Chazal, B.T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman. *Stochastic Convergence of Persistence Landscapes and Silhouettes.* In Journal of Computational Geometry, 6(2), 140-161, 2015] to compute confidence bands for landscapes. As an example compute confidence bands for the expected landscapes for each of the 3 classes in the data set of Exercise 2.

The multiplier bootstrap algorithm.

Input: Landscapes $\lambda_1, \ldots, \lambda_n$; confidence level $1 - \alpha$; number of bootstrap samples B
Output: confidence functions $\ell_n, u_n \colon \mathbb{R} \to \mathbb{R}$

1. Compute the average $\overline{\lambda}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i(t)$, for all t
2. For $j = 1$ to B:
3. Generate $\xi_1, \ldots, \xi_n \sim N(0, 1)$
4. Set $\tilde{\theta}_j = \sup_t n^{-1/2} | \sum_{i=1}^{n} \xi_i \ (\lambda_i(t) - \overline{\lambda}_n(t))|$
5. End for
6. Define $\tilde{Z}(\alpha) = \inf\big\{z : \frac{1}{B} \sum_{j=1}^{B} I(\tilde{\theta}_j > z) \le \alpha\big\}$
7. Set $\ell_n(t) = \overline{\lambda}_n(t) - \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$ and $u_n(t) = \overline{\lambda}_n(t) + \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$
8. Return $\ell_n(t), u_n(t)$