Topology-based Metric Learning

Oliver Gäfvert

KTH Royal Institute of Technology

oliverg@kth.se

Abstract

We present a parametrized class of topology-based metrics that can be optimized over to solve classification problems. The philosophy behind metric learning is that instead of learning a complicated classifier, we learn a metric for which e.g. k-nearest neighbours works well. Using a parametrized class of metrics defined on barcodes we construct a metric learning problem that learns to find a metric that can discriminate between two sets of barcodes.

Introduction

Let $\mathbb{R}_{\geq 0}$ denote the non-negative real numbers and $\mathbb{R}_\infty := \mathbb{R}_{\geq 0} \cup \{\infty\}$. Given the following information:

1. a set of barcodes $B = \{b_1, \dots, b_n\} \subset \text{Mult}(\mathbb{R}^2_{\geq 0})$,
2. a map $\phi \colon B \to \{0, 1\}$ assigning a class to each barcode,

we want to find a (pseudo)metric that discriminates the barcodes of different classes. We will do this through the following construction:

Definition. A **persistence contour** is a function $C \colon \mathbb{R} \times \mathbb{R}_\infty \to \mathbb{R}$ that satisfies:

1. $v \leq C(v, \epsilon) \leq C(v, \tau)$ for $\epsilon \leq \tau$,
2. $C(v, \epsilon) \leq C(w, \epsilon)$ for $v \leq w$,
3. $C(C(v, \epsilon), \tau) \leq C(v, \epsilon + \tau)$.

What we call persistence contour first appeared in [1] under the name superlinear families. Any choice of persistence contour, within some class parametrized by θ, yields a metric on the set of barcodes via the following invariant:

$$\mathfrak{r}_\theta(b)(t) := \sum_{x \in b} \frac{1}{1 + e^{k(C_\theta(x_s, t) - x_e)}}$$

where x_s and x_e denote the start and end of the bar x in b. This is a smoothed version of what is in [2] called the *stable rank function*.

The above invariant lives in $D(\mathbb{R}_{\geq 0})$, the space of decreasing functions from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$, which means that we have a variety of choices of metrics between such invariants. We will consider the interleaving distance. However, to be able to compute the gradient of the metric learning problem, we consider an alternative definition of the interleaving distance defined as follows: the interleaving distance $d_I(f, g)$ between $f, g \in D(\mathbb{R}_{\geq 0})$ is defined as the minimum of the following optimization problem:

$$\begin{aligned} \underset{\epsilon}{\text{minimize}} \quad & \int_0^\infty f(t) - g(t + \epsilon) + g(t) - f(t + \epsilon)\, dt \\ \text{s.t} \quad & f(t) \geq g(t + \epsilon)\ \forall t \in \mathbb{R}_{\geq 0} \\ & g(t) \geq f(t + \epsilon)\ \forall t \in \mathbb{R}_{\geq 0} \end{aligned}$$

Via the above construction, any persistence contour yields a metric on the set of barcodes:

$$d_\theta(b, b') := d_I(\mathfrak{r}_\theta(b), \mathfrak{r}_\theta(b'))$$

A Parametrized Class of Persistence Contours

To be able to describe a class of metrics defined by persistence contours, we need to be able to describe a class of persistence contours. We will consider contours defined by the following equation:

$$t = \int_v^{C_\theta(v,t)} \rho_\theta(\alpha) d\alpha$$

where $\rho_\theta(\alpha)$ is some positive real-valued function. This turns out to yield a very general class of contours. In fact, any continuous and differentiable contour for which condition (3) in the above definition is an equality can be expressed in this way. Note that the class of contours is now determined by what type of parametrized model we choose for ρ_θ. This can be chosen to be:

- a family of polynomials,
- a mixture of Gaussians,
- an artificial neural network,
- a class of step functions.

To be able to optimize over this class of contours however, we need to be able to compute the gradient with respect to θ. For the above equation, this yields for each component:

$$0 = \int_v^{C_\theta(v,t)} \frac{d\rho_\theta(\alpha)}{d\theta_i} d\alpha + \rho_\theta(C_\theta(v, t)) \frac{dC_\theta(v, t)}{d\theta_i}$$

which implies that the ith component of the gradient of $C_\theta(v, t)$ can be expressed as:

$$\Big(\nabla_\theta C_\theta(v, t)\Big)_i = -\frac{1}{\rho_\theta(C_\theta(v, t))} \int_v^{C_\theta(v,t)} \frac{d\rho_\theta(\alpha)}{d\theta_i} d\alpha$$

Metric Learning

To learn a metric for a classification task, we would want to make the elements within each class of barcodes as similar as possible and make elements from different classes as dissimilar as possible. Therefore, the quantity we want to minimize is the following:

$$\underset{\theta}{\text{minimize}} \sum_{b, b' \in \phi^{-1}(0)} d_\theta(b, b') + \sum_{b, b' \in \phi^{-1}(1)} d_\theta(b, b') \tag{1}$$

$$\text{s.t} \sum_{\substack{b \in \phi^{-1}(0) \\ b' \in \phi^{-1}(1)}} d_\theta(b, b') \geq 1 \tag{2}$$

The above optimization problem seems to be non-convex in general, but it depends on the choice of ρ_θ. In any case, we can find a local minima using standard gradient descent methods. Calculating the gradient, at this point, is a straight-forward calculation given a choice of ρ_θ.

An Example

In the following toy example we have that B consists of ten barcodes, with five in each class. We choose $\rho_\theta(\alpha) := \theta_i$ if $i \leq \alpha < i + 1$ and we let $\theta_i = \theta_{i+1}$ if $i \geq 19$. Consequently, $\theta \in \mathbb{R}^{20}$.

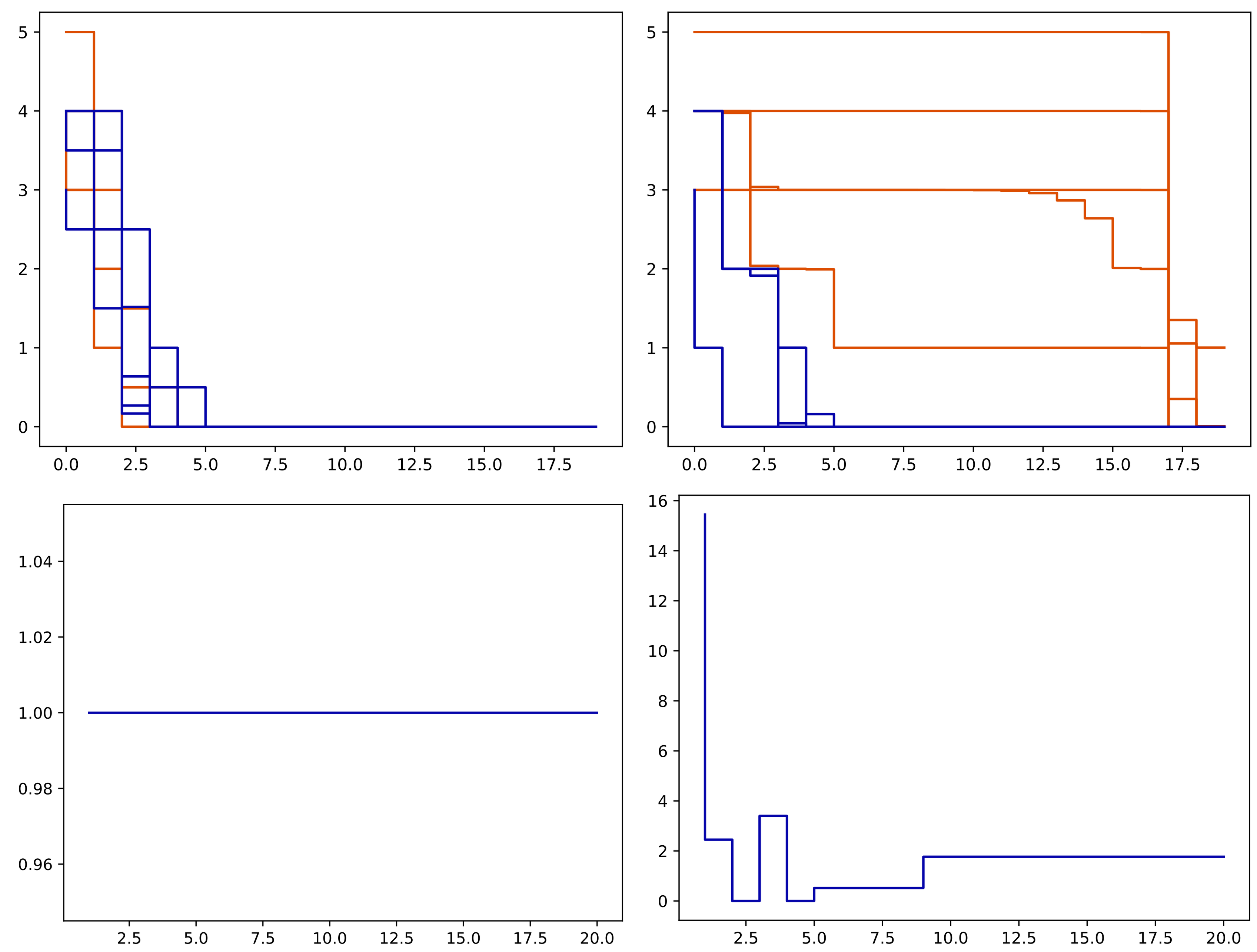

Figure 1: In the upper row we show the curves $\mathfrak{r}_\theta(b)$ for each $b \in B$ where on the left we have that $\theta_i = 1$ for all i (lower left corner), while on the right, θ is as described by the curve in the lower right corner.

References

[1] P. Bubenik, V. de Silva, and J. Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15:1501–1531, 2015.

[2] O. Gäfvert and W. Chachólski. Stable Invariants for Multidimensional Persistence. *ArXiv e-prints*, March 2017.