The optimal assignment kernel is not positive definite
Jean-Philippe Vert

▶ To cite this version:

The optimal assignment kernel is not positive definite

Jean-Philippe Vert
Centre for Computational Biology
Mines ParisTech
Jean-Philippe.Vert@mines.org

January 26, 2008

Abstract

We prove that the optimal assignment kernel, proposed recently as an attempt to embed labeled graphs and more generally tuples of basic data to a Hilbert space, is in fact not always positive definite.

1   Introduction

Let $\mathcal{X}$ be a set, and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a symmetric function that satisfies, for any $n \in \mathbb{N}$ and any $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $(x_1, \ldots, x_n) \in \mathcal{X}^n$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0 \,.$$

Such a function is called a *positive definite kernel* on $\mathcal{X}$. A famous result by [1] states the equivalence between the definition of a positive kernel and the embedding of $\mathcal{X}$ in a Hilbert space, in the sense that k is a positive definite kernel on $\mathcal{X}$ if and only if there exists a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and a mapping $\Phi : \mathcal{X} \to \mathcal{H}$ such that, for any $x, x' \in \mathcal{X}$, it holds that:

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathcal{H}} \,. \tag{1}$$

The construction of positive definite kernels on various sets $\mathcal{X}$ has recently received a lot of attention in statistics and machine learning, because they allow the use of a variety of algorithms for pattern recognition, regression of outlier detection for sets of points in $\mathcal{X}$ [5, 3]. These algorithms,

1

collectively referred to as *kernel methods*, can be thought of as multivariate linear methods that can be performed on the Hilbert space implicitly defined by any positive definite kernel k through (1), because they only access data through inner products, hence through the kernel. This "kernel trick" allows, for example, to perform supervised classification or regression on strings or graphs with state-of-the-art statistical methods as soon as a positive definite kernel for strings or graphs is defined. Unsurprisingly, this has triggered a lot of activity focused on the design of specific positive definite kernels for specific data, such as strings and graphs for applications in bioinformatics in natural language processing [4].

Motivated by applications in computational chemistry, [2] proposed recently a kernel for labeled graphs, and more generally for structured data that can be decomposed into subparts. The kernel, called *optimal assignment kernel*, measures the similarity between two data points by performing an optimal matching between the subparts of both points. It translates a natural notion of similarity between graphs, and can be efficiently computed with the Hungarian algorithm. However, we show below that it is in general not positive definite, which suggests that special care may be needed before using it with kernel methods.

It should be pointed out that not being positive definite is not necessarily a big issue for the use of this kernel in practice. First, it may in fact be positive definite when restricted to a particular set of data used in a practical experiment. Second, other non positive definite kernels, such as the sigmoid kernel, have been shown to be very useful and efficient in combination with kernel methods. Third, practitioners of kernel methods have developed a variety of strategies to limit the possible dysfunction of kernel methods when non positive definite kernels are used, such as projecting the Gram matrix of pairwise kernel values on the set of positive semidefinite matrices before processing it. The good results reported on several chemoinformatics benchmark in [2] indeed confirm the usefulness of the method. Hence our message in this note is certainly not to criticize the use of the optimal assignment kernel in the context of kernel methods. Instead we wish to warn that in some cases, negative eigenvalues may appear in the Gram matrix and specific care may be needed, and simultaneously to contribute to the limitation of error propagation in the scientific litterature.

2 Main result

Let us first define formally the optimal assignment kernel of [2]. We assume given a set $\mathcal{X}'$, endowed with a positive definite kernel k_1 that takes only nonnegative values. The objects we consider are tuples of elements in $\mathcal{X}'$, i.e., an object x decomposes as $x = (x_1, \ldots, x_n)$, where n is the length of the tuple x, denoted $|x|$, and $x_1, \ldots, x_n \in \mathcal{X}'$. We note $\mathcal{X}$ the set of all tuples of elements in $\mathcal{X}'$. Let S_n be the symmetric group, i.e., the set of permutations of n elements. We now recall the kernel on $\mathcal{X}$ proposed in [2]:

Definition 1. *The* optimal assignment kernel $k_A : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is defined, for any $x, y \in \mathcal{X}$, by:*

$$k_A(x, y) = \begin{cases} \max_{\pi \in S_{|x|}} \sum_{i=1}^{|x|} k_1(x_i, y_{\pi(i)}) & \text{if } |y| \geq |x|, \\ \max_{\pi \in S_{|y|}} \sum_{i=1}^{|y|} k_1(x_{\pi(i)}, y_i) & \text{otherwise.} \end{cases}$$

We can now state our main theorem.

Theorem 1. *The optimal alignment kernel is not always positive definite.*

Before proving this results we can make a few comments.

Remark 1. *The meaning of the statement "not always" in Theorem 1 is that there exist choices of $\mathcal{X}'$ and k_1 such that the optimal assignment kernel is positive definite, while there also exist choices for which it is not positive definite.*

Remark 2. *Theorem 1 contradicts Theorem 2.3 in [2], which claims that the optimal assignment kernel is always positive definite. The proof of Theorem 2.3 in [2], however, contains the following error. Using the notations of [2], the author define in the course of their proof the values $A := 2 \sum_{j=1}^{n} v_{n+1} v_j K_{n+1,j}$ and $B := \sum_{j=1}^{n} v_{n+1}^2 K_{n+1,n+1} + v_j^2 K_{jj}$. They show that $A \leq B$, on the one hand, and that $A < 0$, on the other hand. From this they conclude that $B < 0$, which is obvioulsy not a valid logical conclusion.*

In order to prove Theorem 1, we now provide an example of $(\mathcal{X}', k_1)$ pair that leads to a positive definite optimal assignment kernel, and another example that leads to the opposite conclusion.

Lemma 1. *Let $\mathcal{X}' = \{1\}$ be a singleton, and $k_1(1,1) = 1$. Then the optimal assignment kernel is positive definite.*

Proof. When $\mathcal{X}' = \{1\}$, the tuples are simply repeats of the unique element, hence each element $x = (1, \ldots, 1) \in \mathcal{X}$ is uniquely defined by its length $|x| \in \mathbb{N}$. The optimal assignment kernel is then given by:

$$k_A(x, y) = \min(|x|, |y|) .$$

The function $\min(a, b)$ is known to be positive definite on $\mathbb{N}$, therefore k is a valid kernel on $\mathcal{X}$. $\qquad\square$

Lemma 2. *Let $\mathcal{X}' = \mathbb{R}^2$ and $k_1(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$, for $x, y \in \mathbb{R}^2$ and $\gamma > 0$. Then the optimal assignment kernel is not positive definite.*

Proof. The function k_1 defined in Lemma 2 is the well-known Gaussian radial basis function kernel, which is known to be positive definite and only takes nonnegative values, hence it satisfies all hypothesis needed in the definition of the optimal assignment kernel. In order to show that the latter is not positive definite, we exhibit a set of points in $\mathcal{X}$ that can not be embedded in a Hilbert space through (1). For this let us start with four points that form a square in $\mathcal{X}'$, e.g., $A = (0, 0), B = (1, 0), C = (1, 1)$ and $D = (0, 1)$ (Figure 1). Denoting $a := \exp(-\gamma)$, we directly obtain from the definition
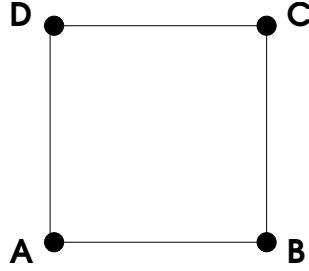

Figure 1: Four points in $\mathcal{X}' = \mathbb{R}^2$, endowed with the positive definite kernel $k_1(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$.

of k_1 that:

$$\begin{cases} k_1(A, A) = k_1(B, B) = k_1(C, C) = k_1(D, D) & = 1 , \\ k_1(A, B) = k_1(B, C) = k_1(C, D) = k_1(D, A) & = a , \\ k_1(A, C) = k_1(B, D) & = a^2 . \end{cases}$$

In the space $\mathcal{X}$ of tuples, let us now consider the six 2-tuples obtained by taking all pairs of distinct points: AB, AC, AD, BC, BD, CD. Using

the definition of the optimal assignment kernel $k(uv, wt) = \max(k_1(u,w) + k_1(v,t), k_1(u,t) + k_1(v,w))$ for $u, v, w, t \in \{A, B, C, D\}$, we easily obtain:

$$\begin{cases} k(AB, AB) = k(AC, AC) = k(AD, AD) = k(BC, BC) \\ \qquad\qquad\qquad\qquad = k(BD, BD) = k(CD, CD) = 2\,, \\ k(AB, AC) = k(AB, BD) = k(BC, BD) = k(BC, AC) \\ \qquad = k(CD, AC) = k(CD, BD) = k(AD, AC) = k(AD, BD) = 1 + a\,, \\ k(AB, BC) = k(BC, CD) = (CD, AD) = k(AB, AD) = 1 + a^2\,, \\ k(AB, CD) = k(AD, BC) = k(AC, BD) = 2a\,. \end{cases}$$

If k was positive definite, then these six 2-tuples could be embedded to a Hilbert space $\mathcal{H}$ by a mapping $\Phi : \mathcal{X} \to \mathcal{H}$ satisfying (1). Let us show that this is impossible. Let $d(x, y) = \|\Phi(x) - \Phi(y)\|_{\mathcal{H}}$ be the Hilbert distance between two points $x, y \in \mathcal{X}$ after their embedding in $\mathcal{H}$. It can be computed from the kernel values by the classical equality:

$$d(x, y)^2 = k(x, x) + k(y, y) - 2k(x, y)\,.$$

We first observe that $d(AB, AC)^2 = d(AC, CD)^2 = 2 - 2a$, and $d(AB, CD)^2 = 4 - 4a$. Therefore,

$$d(AB, CD)^2 = d(AB, AC)^2 + d(AC, CD)^2\,,$$

from which we conclude that (AB, AC, CD) form a half-square, with hypotenuse (AB, CD). A similar computation shows that (AB, BD, CD) is also a half-square with hypotenuse (AB, CD). Moreover,

$$d(AC, BD) = 4 - 4a = d(AB, CD)\,,$$

which shows that the four points (AB, AC, CD, BD) are in fact coplanar and form a square. The same computation when AB and CD are respectively replaced by AD and BC shows that the four points (AD, AC, BC, BD) are also coplanar and also form a square. Hence all six points can be embedded in 3 dimensions, and the points (AB, AD, CD, BC) are themselves coplanar and must form a rectangle on the plane equidistant from AC and BD (Figure 2). The edges of this rectangle have all the same length $d(AB, BC)^2 = d(BC, CD)^2 = d(CD, AD)^2 = d(AD, AB)^2 = 2 - 2a^2$ and is therefore a square, whose hypotenuse (AB, CD) should have a length $\sqrt{4 - 4a^2}$. However a direct computation gives $d(AB, CD) = \sqrt{4 - 4a}$, which provides a contradiction since $0 < a < 1$. Hence the six points can not be embedded into a Hilbert space with k as inner product, which shows that k is not positive definite on $\mathcal{X}$. $\qquad\square$
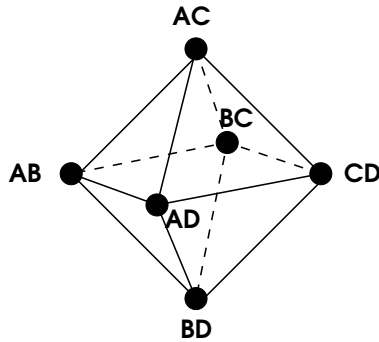
Figure 2: The necessary configuration of the six 2-tuples if embedding with the optimal assignment kernel was possible.

References

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68:337 – 404, 1950.

[2] H. Fröhlich, J. K. Wegner, F. Sieker, and A. Zell. Optimal assignment kernels for attributed molecular graphs. In *Proceedings of the 22nd international conference on Machine learning*, pages 225 – 232, New York, NY, USA, 2005. ACM Press.

[3] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, 2002.

[4] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[5] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New-York, 1998.