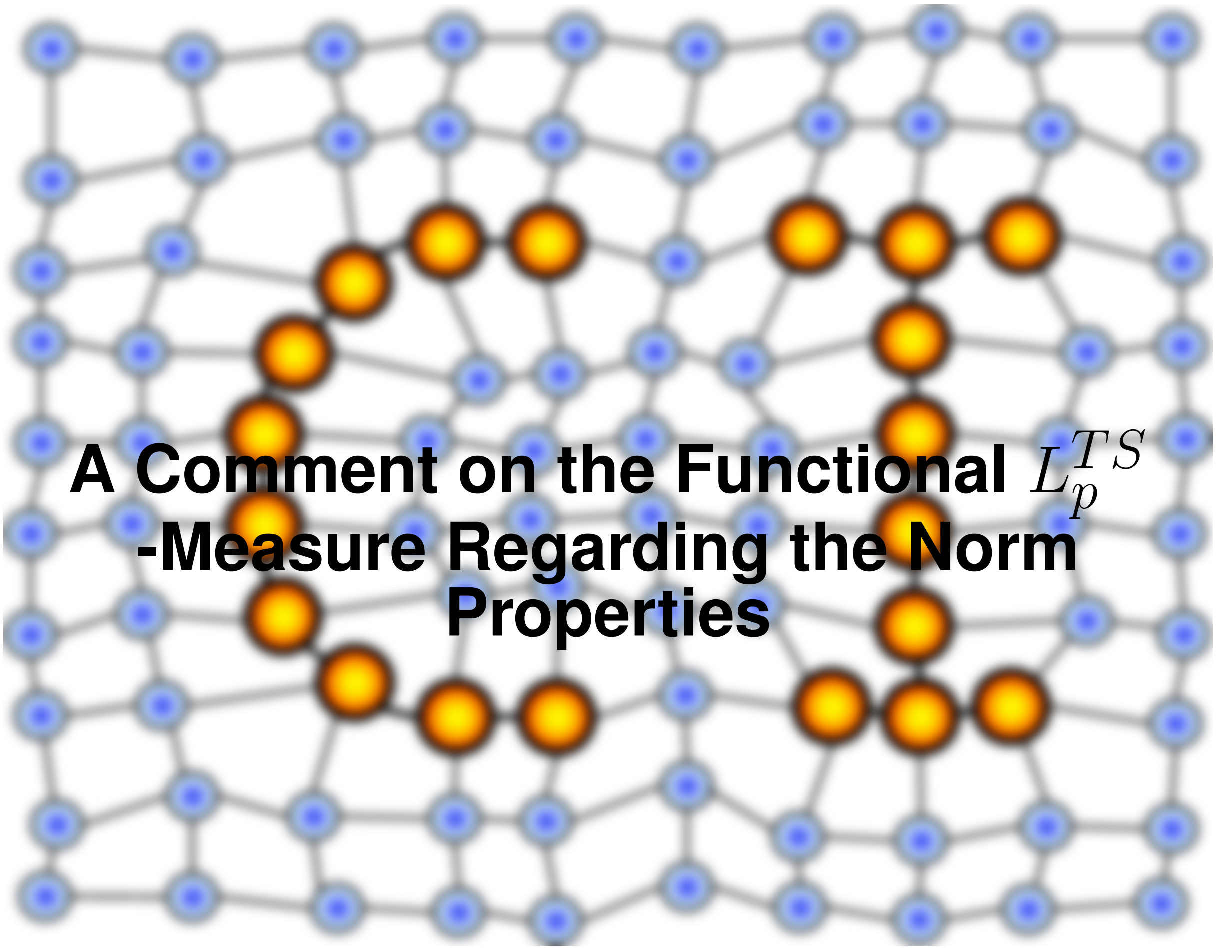
MACHINE LEARNING REPORTS



A Comment on the Functional L_p^{TS} -Measure Regarding the Norm Properties

Report 02/2015
Submitted: 02.02.2015
Published: 04.02.2015

Thomas Villmann[1] and Mandy Lange[1]
(1)University of Applied Sciences Mittweida, Fac. of Mathematics/Natural and Computer Sciences, Computational Intelligence Group, Mittweida, Technikumplatz 17, 09648 Mittweida, Germany

Abstract

In this technical report functional dissimilarity measures are considered. The characteristics of them are investigated and related to those properties particularly assigned to mathematical norms and metrics. One interesting candidate studied in the paper is the functional L T p S measure introduced by Lee and Verleysen.

A Comment on the Functional L_p^{TS}-Measure Regarding the Norm Properties

T. Villmann and M. Lange

Computational Intelligence Group,
University of Applied Sciences Mittweida, Germany

Abstract

In this technical report functional dissimilarity measures are considered. The characteristics of them are investigated and related to those properties particularly assigned to mathematical norms and metrics. One interesting candidate studied in the paper is the functional L_p^{TS} measure introduced by Lee and Verleysen.

1 Introduction - Minkowski norms

Data mining frequently requires the comparison of data vectors. Typically, the (dis-)similarity is judged in terms of a distance, mathematically denoted as a metric. Many, but not all, metrics are generated by norms [6, 13]. Otherwise, for an arbitrary norm $\|\mathbf{v}\|$ for vectors $\mathbf{v}$ of a vector space V, we can always define a metric by

$$d_{\|\bullet\|}(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1 - \mathbf{v}_2\| \ . \tag{1}$$

Yet, frequently weaker dissimilarity measures than metrics are considered in machine learning [1, 4, 5, 17]. For example, the squared Euclidean distance, usually applied in neural networks, violates the triangle inequality, which is required to be fulfilled by a metric. Hence, it is only a *quasi-metric* following the categorization system suggested by PEKALSKA&DUIN in [11]. Accordingly, we denote a measure

$$m(\mathbf{v}) : V \longrightarrow \mathbb{R}^{\oplus} \tag{2}$$

with $\mathbb{R}^{\oplus}$ being the set of non-negative real numbers, as a *quasi-norm*, if all norm properties are satisfied except the triangle inequality. Obviously, quasi-norms generate by means of (1) quasi-metrics at least.

The l_p-norm for vectors $\mathbf{x} \in \mathbb{C}^n$ defined as

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{k=1}^{n} |x_k|^p} \tag{3}$$

with $1 \leq p \leq \infty$ is a standard norm in data mining [8]. It is also denoted as the Minkowski norm defining the Minkowski metric $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p$. Note that if $0 < p < 1$ is valid, $\|\mathbf{x}\|_p$ is only a quasi-norm whereas $(d_p(\mathbf{x}, \mathbf{y}))^p = \|\mathbf{x} - \mathbf{y}\|_p^p$ generates a metric [2].

For the Euclidean case $p = 2$, the pair $\left(\mathbb{C}^n, \|\bullet\|_p\right)$ is a Hilbert space equipped with the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_E = \sum_{k=1}^{n} x_k \bar{y}_k \tag{4}$$

where $\bar{y}_k$ is the conjugate complex of y_k. For $p \neq 2$, $\left(\mathbb{C}, \|\bullet\|_p\right)$ forms only a Banach space. For Banach spaces weaker counterparts to inner products exists.

These are so-called semi-inner products (SIP, [10]), which are generally not unique but also generating the norm.[1] Yet, for the l_p-norm it is uniquely given as

$$[\mathbf{x},\mathbf{y}]_p = \frac{1}{\left(\|\mathbf{y}\|_p\right)^{p-2}} \sum_{i=1}^{n} x_i \cdot \bar{y}_i \cdot |y_i|^{p-2} \tag{9}$$

which becomes

$$[\mathbf{x},\mathbf{y}]_p = \frac{1}{\left(\|\mathbf{y}\|_p\right)^{p-2}} \sum_{i=1}^{n} x_i \, |y_i|^{p-1} \, sgn\,(y_i)$$

in case of real valued vectors. Here

$$sgn\,(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} . \tag{10}$$

is the signum function.

The continuity of the SIP as a map

$$[\mathbf{x},\mathbf{y}]_p : \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C} :$$

[1] A semi-inner product (SIP) $[\bullet,\bullet]$ of a general vector space V is a map

$$[\bullet,\bullet] : V \times V \longrightarrow \mathbb{C} \tag{5}$$

with the following properties:

1. positive semi-definite

   $$[\mathbf{x},\mathbf{x}] \geq 0 \tag{6}$$

   and $[\mathbf{x},\mathbf{x}] = 0$ iff $\mathbf{x} = \mathbf{0}$

2. linear with respect to the first argument for $\xi \in \mathbb{C}$

   $$\xi \cdot [\mathbf{x},\mathbf{z}] + [\mathbf{y},\mathbf{z}] = [\xi \cdot \mathbf{x} + \mathbf{y},\mathbf{z}] \tag{7}$$

3. Cauchy-Schwarz inequality

   $$|[\mathbf{x},\mathbf{y}]|^2 \leq [\mathbf{x},\mathbf{x}]\,[\mathbf{y},\mathbf{y}] \tag{8}$$

We emphasize that, in contradiction to inner products, SIPs may violate the symmetry condition, i.e. we generally have $[\mathbf{x},\mathbf{y}] \neq \overline{[\mathbf{y},\mathbf{x}]}$. However, each SIP generates a norm via $\|\mathbf{x}\| = \sqrt{[\mathbf{x},\mathbf{x}]}$. Vice versa, each norm corresponds to a SIP, which, however, not need to be unique. For further reading we refer to [3, 10].

can be related to the differentiability of the respective norm [3].

Accordingly, the Banach spaces $\widehat{\mathcal{L}}_p$ of complex Lebesgue-integrable functions are equipped with the respective SIP

$$[f, g]_p = \frac{1}{\left(\|g\|_p\right)^{p-2}} \int f \cdot \bar{g} \cdot |g|^{p-2}\, dt \tag{11}$$

for complex functions g and f and

$$\|f\|_p = \sqrt{[f, f]_p} \tag{12}$$

In case of the space $\mathcal{L}_p$ of real Lebesgue-integrable functions we have

$$[f, g]_p = \frac{1}{\left(\|g\|_p\right)^{p-2}} \int f \cdot |g|^{p-1} \cdot sgn\,(g\,(t))\, dt \tag{13}$$

as SIP. Again, for $p = 2$ both spaces are Hilbert spaces with the usual inner product.

Although $\mathcal{L}_p$ and $\widehat{\mathcal{L}}_p$ are function spaces, the respective norms as well as SIPs do not explicitly take the functional character into account, i.e. the norm values are invariant under such transformations of the function, which switch the function values for two arbitrary arguments t_1 and t_2. The same property we observe for the discrete versions according to the switch of vector dimensions. For this reasons so-called functional norms come into play [12].

2 Functional norms

2.1 The Sobolev norm

As explained above, the Minkowski-norm is not a functional norm, i.e. it does not explicitly take into account the functional character. One of the most prominent functional norms is the *Sobolev-norm*

$$\|f\|_{K,p} = \left[\sum_{|\alpha| \leq K} \left(\|D^\alpha f\|_p\right)^p\right]^{\frac{1}{p}} \tag{14}$$

assigned to the *Sobolev-space* $\mathcal{W}_{K,p} = \{f | D^\alpha f \in \mathcal{L}_p, |\alpha| \leq K\}$ of (real) differentiable functions up to order K where $D^\alpha = \frac{\partial^{|\alpha|}}{\partial \alpha_1 \ldots \partial \alpha_{|\alpha|}}$ is the differential operator.

It is well-known that $\mathcal{W}_{K,p}$ is a Hilbert spaces only for $p = 2$ as $\mathcal{L}_p$ does. The related SIP

$$[f, g]_{K,p} = \frac{1}{\|g\|_{K,p}^{p-2}} \sum_{|\alpha| \leq K} \int f^{(\alpha)} \cdot \left|g^{(\alpha)}\right|^{p-1} sgn\left(g^{(\alpha)}\right) dt$$

is similar to (13) for $\mathcal{L}_p$ [7]. We observe that the Sobolev-norm requires the differentiability of the functions, which might be a disadvantage, if this property cannot be ensured.

However, frequently only discrete approximations of functions are considered in data mining, i.e. vectors $\mathbf{x} \in \mathbb{R}^n$ are discrete representations of functions. Thus, vector entries x_k and x_{k+j} are functionally related depending on the index difference j. Obviously, the l_p-norm (3) does not make use of these relations. Otherwise, machine learning algorithms in data mining may benefit from those functional data properties [14, 15, 16].

2.2 The functional measure L_p^{TS} from Lee and Verleysen

An attempt to generate an alternative discrete functional norm was made by Lee&Verleysen in [9]. To this extend, they considered time series x_k collected in a vector $\mathbf{x} = (x_0, \ldots, x_{D+1}) \in \mathbb{R}^n$ with $n = D+2$ and introduced their dissimilarity measure

$$\delta_p(\mathbf{x}, \mathbf{y}) = L_p^{TS}(\mathbf{x} - \mathbf{y}, \tau) \tag{15}$$

with

$$L_p^{TS}(\mathbf{x}, \tau) = \left(\sum_{i=1}^{D} (g_i(\mathbf{x}, \tau))^p\right)^{\frac{1}{p}} \tag{16}$$

where the local function $g_i(\mathbf{x}, \tau)$ is the sum

$$g_i(\mathbf{x}, \tau) = A_i(\mathbf{x}, \tau) + B_i(\mathbf{x}, \tau) \tag{17}$$

of the areas A_i and B_i assigned to the left and right sides of x_i, respectively, as depicted in Fig.1.

The parameter τ can be interpreted as sampling period if $\mathbf{x}$ represents a time series $x_i = x(t_i)$ of discrete time points t_i. The areas A_i and B_i are calculated case-dependently as

$$A_i(\mathbf{x}, \tau) = \begin{cases} \frac{\tau}{2}|x_i| & 0 \leq x_i x_{i-1} \\ \frac{\tau}{2}\frac{|x_i|^2}{|x_i| + |x_{i-1}|} & 0 > x_i x_{i-1} \end{cases} \tag{18}$$

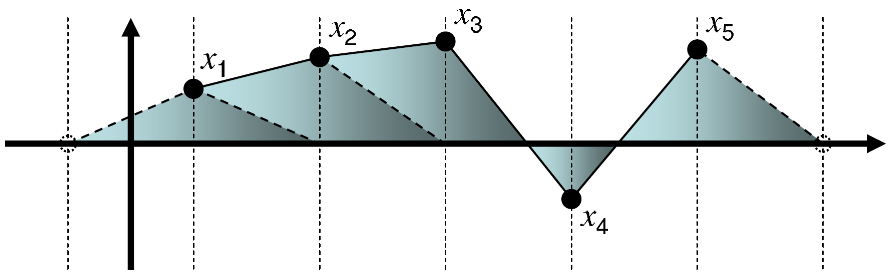


Figure 1: Visualization of the quasi-norm L_p^{TS} taken from [9]. This quasi-norm involves the areas of the triangles placed on the left (A_i) and right (B_i) side of each coordinate.

and

$$B_i(\mathbf{x},\tau) = \begin{cases} \frac{\tau}{2}|x_i| & 0 \le x_i x_{i+1} \\ \frac{\tau}{2}\frac{|x_i|^2}{|x_i|+|x_{i+1}|} & 0 > x_i x_{i+1} \end{cases} \tag{19}$$

respectively. Further it is assumed that $x_0 = x_{D+1} = 0$ is valid. The quantity $L_p^{TS}(\mathbf{x},\tau)$ defined in (16) was proposed in [9] to be a norm. However, this statement has to be corrected. One can easily verify that the triangle inequality may be violated.[2] Hence, $L_p^{TS}(\mathbf{x},\tau)$ is only a quasi-norm and, consequently, the quantity $\delta_p(\mathbf{x},\mathbf{y})$ from (15) is only a quasi-metric.

In the next step we briefly study the approximation properties of the quasi-norm $L_p^{TS}(\mathbf{x},\tau)$ with respect to the function norm

$$L_p[x(t)] = \sqrt[p]{\int |x(t)|^p \, dt} \, . \tag{20}$$

For this purpose, we consider, as suggested when introduced in [9], the vector $\mathbf{x}$ to be a discrete representation of a continuous function $x(t)$. Then the difference $i-(i-1)$ corresponds to a small interval Δt scaled by sampling period τ in the model. We consider the function

$$\alpha(x(t),t,\tau,\Delta t) = \frac{\tau}{2}|x(t)| \cdot \left(H(x(t)\cdot x(t-\Delta t)) + \frac{1-H(x(t)\cdot x(t-\Delta t))}{1+\frac{|x(t-\Delta t)|}{|x(t)|}} \right) \tag{21}$$

[2]Consider the following example: $\mathbf{x} = (0,10,1,10,1,0)^T$, $\mathbf{y} = (0,10,-1,10,-1,0)^T$, which yield $\mathbf{x}+\mathbf{y} = (0,20,0,20,0,0)^T$. For $p=2$ and $\tau=1$ we obtain $L_p^{TS}(\mathbf{x},\tau) \approx 14.21$, $L_p^{TS}(\mathbf{y},\tau) \approx 13.19$ and $L_p^{TS}(\mathbf{x}+\mathbf{y},\tau) \approx 28.28$. Thus $L_p^{TS}(\mathbf{x}+\mathbf{y},\tau) > L_p^{TS}(\mathbf{x},\tau) + L_p^{TS}(\mathbf{y},\tau)$ is obtained, violating the triangle inequality.

playing the role of the functional counterpart of $A_i(\mathbf{x},\tau)$. Here

$$H(z)=\begin{cases}1 & ;\ z\geq 0\\ 0 & ;\ z<0\end{cases}$$

is the Heaviside function. Analogously, we have

$$\beta(x(t),t,\tau,\Delta t)=\frac{\tau}{2}|x(t)|\cdot\left(H(x(t)\cdot x(t+\Delta t))+\frac{1-H(x(t)\cdot x(t+\Delta t))}{1+\frac{|x(t+\Delta t)|}{|x(t)|}}\right)$$

as the functional complement to $B_i(\mathbf{x},\tau)$. Summation of both yields

$$\gamma(x(t),t,\tau,\Delta t)=\alpha(x(t),t,\tau,\Delta t)+\beta(x(t),t,\tau,\Delta t) \tag{22}$$

as a $\tau\cdot\Delta t$-dependent counterpart of the local function $g_i(\mathbf{x},\tau)$ from (17) for the discrete case. Interpreting the term

$$\begin{aligned}\vartheta_x(t,\tau,\Delta t)=&H(x(t)\cdot x(t-\Delta t))+H(x(t)\cdot x(t+\Delta t))\\&+\frac{1-H(x(t)\cdot x(t-\Delta t))}{1+\frac{|x(t-\Delta t)|}{|x(t)|}}+\frac{1-H(x(t)\cdot x(t+\Delta t))}{1+\frac{|x(t+\Delta t)|}{|x(t)|}}\end{aligned} \tag{23}$$

as a multiplicative deviation of $x(t)$ we rewrite the local function (22) as

$$\gamma(x(t),t,\tau,\Delta t)=\frac{\tau}{2}|x(t)\cdot\vartheta_x(t,\tau,\Delta t)| \tag{24}$$

where we made use of the observation that $\vartheta_x(t,\tau,\Delta t)\geq 0$ is valid.

The deviation function $\vartheta_x(t,\tau,\Delta t)$ is not necessarily continuous everywhere with respect to the difference Δt, although $x(t)$ is assumed to be a continuous function. To see this, we suppose a continuous time-dependent function $x(t)$ on the interval $[a,b]$ with $x(a)\cdot x(b)<0$. We suppose w.l.o.g. $x(a)<0$. Then exists at least one t_0 with $x(t_0)=0$ together with an $\varepsilon>0$ determining the interval $I_\varepsilon(t_0)=[t_0-\varepsilon,t_0+\varepsilon]$ such that the following statements hold:

1. $I_\varepsilon(t_0)\subseteq[a,b]$,
2. $x(t_0-\varepsilon)\cdot x(t_0+\varepsilon)<0$,
3. $x(t)$ is monotonically increasing in $I_\varepsilon(t_0)$,
4. $x(t)<0$ for $t\in[t_0-\varepsilon,t_0)$ and $x(t)>0$ for $t\in(t_0,t_0+\varepsilon]$.

Let $t^* \in (t_0 - \varepsilon, t_0)$ arbitrarily but fixed and $\Delta t = t - t^* < \frac{\varepsilon}{2}$. We recognize the strong inequality $x(t_0) > x(t^*)$ and consider the limit $t^* \to t$. In this case we obtain

$$\lim_{\Delta t \searrow 0} \alpha\left(x\left(t_0\right), t_0, \tau, \Delta t\right) = \lim_{\Delta t \searrow 0} \frac{\tau}{2} \cdot \frac{\left|x\left(t_0\right)\right|}{1 + \frac{\left|x\left(t_0 - \Delta t\right)\right|}{\left|x\left(t_0\right)\right|}} = \frac{\tau}{4} \cdot \left|x\left(t_0\right)\right| \neq \alpha\left(x\left(t_0\right), t_0, \tau, 0\right)$$

for the function $\alpha\left(x\left(t\right), t, \tau, \Delta t\right)$ from (21) because of $\alpha\left(x\left(t_0\right), t_0, \tau, 0\right) = \frac{\tau}{2} \cdot \left|x\left(t_0\right)\right|$ is valid. Analogously, the function term $\beta\left(x\left(t\right), t, \tau, \Delta t\right)$ can be treated. Hence, the deviation $\vartheta_x\left(t, \tau, \Delta t\right)$ in (24) is not necessarily continuous and, therefore, we can not obtain a continuous approximation of the function norm L_p from (20) by means of $L_p^{TS}\left(\mathbf{x}, \tau\right)$.

3 Conclusion

In this paper we considered dissimilarities for functional data. We briefly discussed their properties. One particular focus was the functional measure L_p^{TS}. In contradiction to earlier statements, it violates the triangle inequality and, therefore, is only a quasi-norm.

References

[1] C. Bishop. *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC, New York, NY, 2006.

[2] M. Day. The spaces L^p with $0 < p < 1$. *Bulletin of the American Mathematical Society*, 46:816–823, 1940.

[3] J. Giles. Classes of semi-inner-product spaces. *Transactions of the American Mathematical Society*, 129:436–446, 1967.

[4] S. Haykin. *Neural Networks - A Comprehensive Foundation.* IEEE Press, New York, 1994.

[5] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences.* Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[6] A. Kolmogorov and S. Fomin. *Reelle Funktionen und Funktionalanalysis.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1975.

[7] M. Lange, M. Biehl, and T. Villmann. Non-Euclidean principal component analysis by Hebbian learning. *Neurocomputing*, 147:107–119, 2015.

[8] M. Lange, D. Zühlke, O. Holz, and T. Villmann. Applications of l_p-norms and their smooth approximations for gradient based learning vector quantization. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2014)*, pages 271–276, Louvain-La-Neuve, Belgium, 2014. i6doc.com.

[9] J. Lee and M. Verleysen. Generalization of the l_p norm for time series and its application to self-organizing maps. In M. Cottrell, editor, *Proc. of Workshop on Self-Organizing Maps (WSOM) 2005*, pages 733–740, Paris, Sorbonne, 2005.

[10] G. Lumer. Semi-inner-product spaces. *Transactions of the American Mathematical Society*, 100:29–43, 1961.

[11] E. Pekalska and R. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications.* World Scientific, 2006.

[12] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer Science+Media, New York, 2nd edition, 2006.

[13] F. Riesz and B. Sz.-Nagy. *Vorlesungen über Functionalanalysis*. Verlag Harri Deutsch, Frankfurt/M., 4th edition, 1982.

[14] F. Rossi, B. Conan-Guez, and A. E. Golli. Clustering functional data with the SOM algorithm. In M.Verleysen, editor, *European Symposium on Artificial Neural Networks 2004*, pages 305–312. d-side publications, 2004.

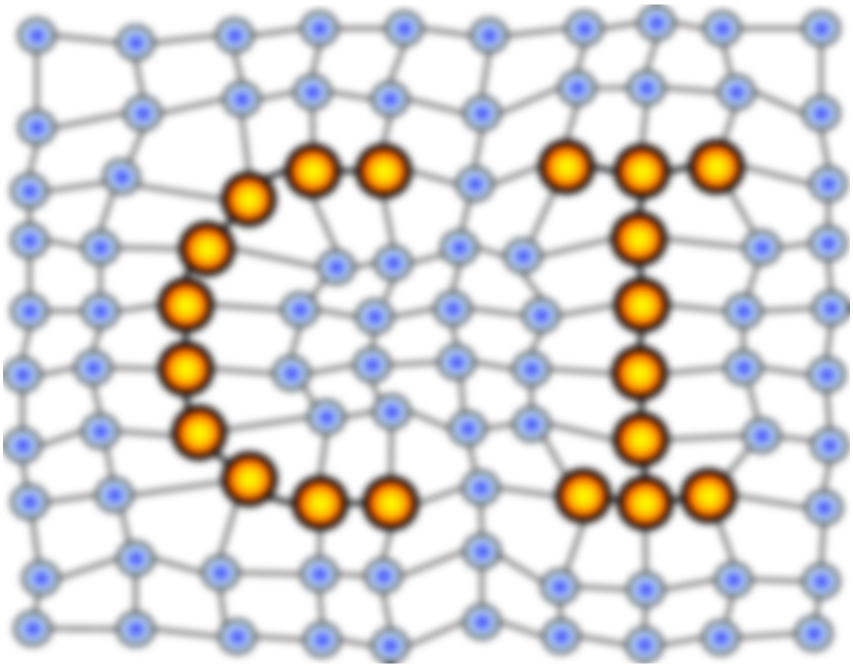
[15] F. Rossi, N. Delannay, B. Conan-Gueza, and M. Verleysen. Representation of functional data in neural networks. *Neurocomputing*, 64:183–210, 2005.

[16] T. Villmann, M. Kästner, D. Nebel, and M. Riedel. Lateral enhancement in adaptative metric learning for functional data. *Neurocomputing*, 131:23–31, 2014.

[17] T. Villmann, F.-M. Schleif, M. Kaden, and M. Lange, editors. *Advances in Self-Organizing Maps and Learning Vector Quantization - Proceedings of the 10th International Workshop, WSOM 2014, Mittweida, Germany*. Number 295 in Advances in Telligent Systems and Computing. Springer, Heidelberg, 2014.

Machine Learning Reports

Report 02/2015



Impressum

Machine Learning Reports ISSN: 1865-3960

▽ Publisher/Editors

Prof. Dr. rer. nat. Thomas Villmann
University of Applied Sciences Mittweida
Technikumplatz 17, 09648 Mittweida, Germany
• http://www.mni.hs-mittweida.de/

Dr. rer. nat. Frank-Michael Schleif
University of Bielefeld
Universitätsstrasse 21-23, 33615 Bielefeld, Germany
• http://www.cit-ec.de/tcs/about

▽ Copyright & Licence

▽ Acknowledgments

We would like to thank the reviewers for their time and patience.