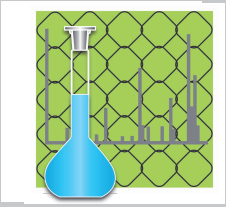
doi:10.2533/chimia.2026.720 Chimia 80 (2026) 720 © Swiss Chemical Society



# Highlights of Analytical Sciences in Switzerland

## Division of Analytical Sciences

A Division of the Swiss Chemical Society

## The Infinite-Dimensional Nature of Spectroscopy and Why Models Succeed, Fail, and Mislead

Umberto Michelucci[a] and Francesca Venturini*[,b]

*Correspondence: F. Venturini, E-mail: vent@zhaw.ch
[a]HSLU Lucerne University of Applied Sciences and Arts, Computer Science Department, CH-6343 Risch-Rotkreuz, Switzerland; [b]ZHAW Zurich University of Applied Sciences, School of Engineering, Institute of Applied Mathematics and Physics, Technikumstr. 9, CH-8401 Winterthur, Switzerland.

Keywords: Classification · Infinite dimensions · Machine learning · Spectroscopy

In his famous book, '*Flatland: A Romance of Many Dimensions*', Abbott describes a two-dimensional world. Given the limitations of their world, inhabitants of Flatland cannot do many things that are easy for us. For example, consider a typical Flatland inhabitant: Joe. If you put a fence without gates around Joe, he will never be able to escape (remember he can only move within the two-dimensional world). But if you alter the geometry of the world, adding a third dimension, suddenly Joe can hop over the fence and go on his merry way.

Why is this relevant for spectroscopy? Our research shows that something similar happens when we have a classification problem in machine learning. You can see an example of how dimensions influence classification in Fig. 1.

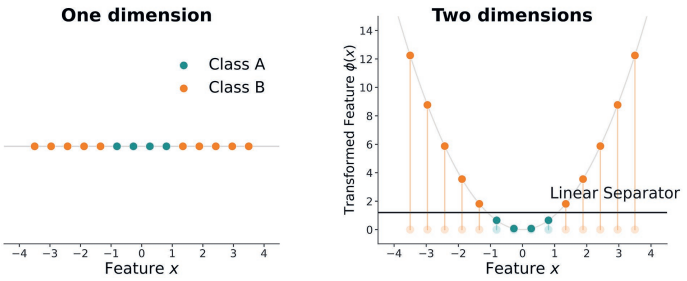



Fig. 1. In one dimension it is impossible to classify the data with a linear separator. By adding one dimension classification becomes possible.

Now, consider two one-dimensional Gaussians with the same mean and a slightly different standard deviation. Given a point sampled from one, it is not possible to determine which Gaussian it comes from, or in other words, to classify it. The reason is that the two probability distributions overlap significantly. Now let's change the geometry of the world in which the Gaussians (our Joe) live. If we increase the number of dimensions of the world, something incredible happens. In the limit of infinite dimensions (admittedly more than the three that Joe needs), it suddenly becomes possible to tell if a point belongs to one Gaussian or the other with absolute certainty (see Fig. 2). The additional dimensions give us a way of hopping over the fence of inseparability. Increasingly high dimensionality brings Gaussians (our Joe) in a 'world' where a perfect classification becomes not only possible, but increasingly easy.

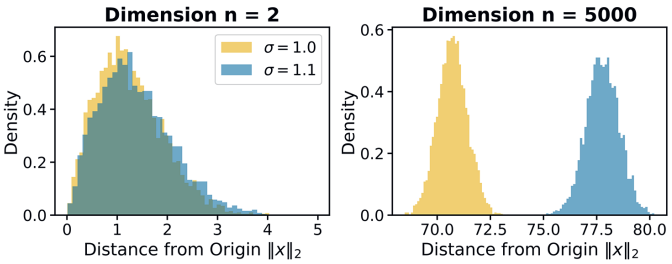



Fig. 2. Increasing the number of dimensions n makes two Gaussian distributions with same mean but slightly different standard deviation σ easily separable as all values concentrates on the outer part of the distribution.

Spectra are characterized by a number of intensities of the order of thousands (*e.g.* 1024, 2048), depending on the instrumentation. Thus, spectra can be considered to 'exist' in a very high-dimensional world. One clearly needs to find the right algorithm (Joe needs to think that he can hop over the fence). But our research shows that the incredible success in classification in spectroscopy is often explained by the intrinsic high dimensionality of spectra.

Thus, it is important for researchers to check if the good accuracy of their model is due to the algorithm really learning the chemistry underlying the process or due to the geometry of space. To make this distinction clear, imagine we put a fence around Joe but this time with a gate. We would like Joe to learn that the gate allows him to go in and out of the fenced area without disappearing from his 2D world and, thus, scaring his fellow inhabitants. But Joe might simply see hopping over the fence the easiest solution. He might not even know what a gate is. The same can happen with algorithms. They will look for the fastest way of solving a problem, and a favorable space geometry is an efficient fantastic shortcut. The consequence is that the success of the models is not transferable to other instruments, samples, problems because they rely on specific statistical characteristics of the data.

In conclusion, spectra live in a very high-dimensional space with a geometry radically different from what our intuition can understand. To solve a classification task in such a space, algorithms can find shortcuts that are completely detached from the chemistry of the underlying processes and substances. In our paper we show these shortcuts, discuss how they appear when doing machine learning on spectra, and discuss how to address this challenge. Our research focusses on understanding and explaining how to apply machine learning to spectroscopy, so that models learn from underlying chemical processes and keep into account the 'world' in which spectra exist.[1] **We hope that our contribution might shed some light on how hopping fences might be an easy way out for algorithms.**

Received: September 10, 2026

[1] U. Michelucci, F. Venturini, *Analyst* **2026**, *151*, 2978, https://doi.org/10.1039/D6AN00346J.

**Can you show us your analytical highlight?**
Please contact: Dr. Bodo Hattendorf, ETH Zurich, HCI G105, Vladimir-Prelog-Weg 1, CH-8093 Zurich, Tel.: +41 44 632 44 72
E-mail: analytical_highlights@chimia.ch