Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Cluster-Based Classification with Neural ODEs via Control

Antonio Álvarez-López * [1,3], Rafael Orive-Illera † [1,2], and Enrique Zuazua ‡ [1,3,4]

[1]Departamento de Matemáticas, Universidad Autónoma de Madrid, Madrid 2804, Spain.
[2]Instituto de Ciencias Matemáticas, CSIC-UAM-UC3M-UCM, Madrid 2804, Spain.
[3]Chair for Dynamics, Control, Machine Learning, and Numerics, Alexander von Humboldt-Professorship, Department of Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen 91058, Germany.
[4]Chair of Computational Mathematics, Fundación Deusto, Bilbao 48007, Spain.

Abstract. We address binary classification using neural ordinary differential equations from the perspective of simultaneous control of N data points. We consider a single-neuron architecture with parameters fixed as piecewise constant functions of time. In this setting, the model complexity can be quantified by the number of control switches. Previous work has shown that classification can be achieved using a point-by-point strategy that requires $\mathcal{O}(N)$ switches. We propose a new control method that classifies any arbitrary dataset by sequentially steering clusters of d points, thereby reducing the complexity to $\mathcal{O}(N/d)$ switches. The optimality of this result, particularly in high dimensions, is supported by some numerical experiments. Our complexity bound is sufficient but often conservative because same-class points tend to appear in larger clusters, simplifying classification. This motivates studying the probability distribution of the number of switches required. We introduce a simple control method that imposes a collinearity constraint on the parameters, and analyze a worst-case scenario where both classes have the same size and all points are i.i.d. Our results highlight the benefits of high-dimensional spaces, showing that classification using constant controls becomes more probable as d increases.

Keywords:
Classification,
Complexity,
Neural ODEs,
Separability,
Controllability.

Article Info.:
Volume: 4
Number: 2
Pages: 128 - 156
Date: June/2025
doi.org/10.4208/jml.241114

Article History:
Received: 14/11/2024
Accepted: 13/04/2025

Communicated by:
Qianxiao Li

1 Introduction

At the heart of machine learning lies supervised learning [36], a framework that has been successfully applied in a vast number of domains [4, 37]. The main objective is to learn an unknown mapping $F : \mathscr{X} \to \mathscr{Y}$. To achieve this, a model $\hat{F} : \mathscr{X} \to \mathscr{Y}$ to approximate F is constructed by minimizing a loss function, using only the available – possibly noisy – values of F over a finite dataset $\mathscr{D} \subset \mathscr{X} \times \mathscr{Y}$. Our focus is on evaluating the minimal complexity required for $\hat{F}$ to fit the points in $\mathscr{D}$ without error.

In the context of data classification, the range of F is finite, and its elements are referred to as labels. Over the years, a wide variety of models have been developed, with notable examples including linear discriminants [21], support vector machines [10], random

*Corresponding author. antonio.alvarezl@uam.es
†rafael.orive@icmat.es
‡enrique.zuazua@fau.de

forests [7], and neural networks [45]. In [16], a methodology from a control perspective is proposed, based on modeling deep residual networks (ResNets, [28]) as continuous-time dynamical systems known as neural ordinary differential equations (neural ODEs).

Neural ODEs have seen the development of several variants [8,26,41], yet the standard form remains as

$$\begin{cases} \dot{\mathbf{x}}(t) = W(t)\,\sigma\big(A(t)\,\mathbf{x}(t) + \mathbf{b}(t)\big), & t \in (0,T), \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \tag{1.1}$$

where

- $\mathbf{x}_0 \in \mathbb{R}^d$ is an input point.
- $(W, A, \mathbf{b}) \in L^\infty((0,T), \mathbb{R}^{d\times p} \times \mathbb{R}^{p\times d} \times \mathbb{R}^p)$ are parameters to be trained.
- $d, p \geq 1$ are the state dimension and the width of the model, respectively.
- $\sigma : \mathbb{R}^p \to \mathbb{R}^p$ is a prefixed nonlinear Lipschitz function applied component-wise.

Existence and uniqueness of the solution to (1.1) is guaranteed by the Cauchy-Lipschitz theorem, ensuring the well-definedness of the flow map

$$\Phi_t(\cdot; W, A, \mathbf{b}) : \mathbf{x}_0 \in \mathbb{R}^d \quad \longmapsto \quad \mathbf{x}(t) \in \mathbb{R}^d, \quad t \in [0,T]. \tag{1.2}$$

The formulation of (1.1) naturally frames supervised learning as a control problem. Here, the input space is $\mathscr{X} = \mathbb{R}^d$, and the parameters $(W, A, \mathbf{b})$ serve as controls that simultaneously guide all input points toward their respective target positions in $\mathbb{R}^d$. To match the output space, a mapping $g : \mathbb{R}^d \to \mathscr{Y}$ is introduced as a final layer. The complete model is thus defined by the composition $\hat{F} = g \circ \Phi_T$.

We focus on binary classification with a hard classifier, whereby $\mathscr{Y} = \{1, 0\}$ and g is the characteristic function of a fixed set. Nonetheless, our results can be extended to any multiclass setting by fixing g as a weighted sum of predefined characteristic functions, each corresponding to a distinct label.

Neural ODEs were originally conceived as a tool for understanding deep ResNets, but they have since made a significant impact on machine learning. Their continuous-time framework facilitates mathematical analysis and provides practical benefits like incorporating structure or the design of new discrete schemes. For more details, we refer to [16].

Notation

- Scalars are denoted by plain letters, vectors by boldface letters, and matrices by uppercase letters. The scalar product of two vectors $\mathbf{u}, \mathbf{v}$ is written as $\mathbf{u} \cdot \mathbf{v}$.
- Subscripts identify elements of a set. Superscripts identify components of a vector.
- $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ denotes the canonical basis in $\mathbb{R}^d$.
- $\mathbb{S}^{d-1}$ denotes the $(d-1)$-dimensional sphere in $\mathbb{R}^d$.
- The cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$.
- For $x \in \mathbb{R}$, we write $\lceil x \rceil := \min\{n \in \mathbb{Z} : n \geq x\}$, $\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \leq x\}$.