Functorial wrappers for high-dimensional classification algorithms

James Carzon*

April 1, 2022

Abstract

Technology has been increasingly increasing in recent years. Previous literature on solving real-life classification problems uses standard non-parametric regression algorithms such as nearest neighbor methods which learn *a* true regression function perfectly well as the sample size goes to infinity due to their model flexibility. Then a classifier may be constructed by binning the regression outputs. In high-dimensional settings, these algorithms struggle to estimate the truth well due to the curse of dimensionality. In this work, we provide a robust and efficient way to handle high-dimensional classification problems and exploit the algebraic structure underlying most real life data sets by wrapping algorithms with functors which preserve only important covariate group structure when mapping to the response group. We provide some numerical results.

1 Introduction

Let $\mathcal{X}$ be a vector space homomorphic to $\mathbb{R}^{d_1} \otimes \mathbb{Q}^{d_2} \otimes \mathbb{Z}^{d_3} \otimes \mathbb{G}^{d_4} \otimes \mathbb{Z}_{k_1} \otimes \cdots \otimes \mathbb{Z}_{k_\ell}$, where $\mathbb{R}$ denotes the real numbers, $\mathbb{Q}$ denotes the rationals, $\mathbb{Z}$ denotes the integers (and $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$ as usual). Let $\mathcal{Y} \subset \mathbb{R}^{d_5}$ be the range of outcomes from $\mathcal{X}$ under some unknown continuous, bounded map $f : \mathcal{X} \to \mathbb{R}$. We are interested in inferring the map f from some random observations $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n) \in \mathbb{P}_{\mathcal{X}\times\mathcal{Y}}$, where the distribution $\mathbb{P}$ is unknown. One way to do that is to pass the observed covariates X_i through a *forgetful functor* as is used in category theory.

Definition. Let $\mathcal{C}$ be a category based on sets. We say a functor F is a *forgetful functor* from $\mathcal{C}$ to another category $\mathcal{D}$ if for every morphism $\mu \in \{D \to D\}$ there exists a morphism $\lambda \in \{C \to C\}$ such that

$$F(\lambda) = \mu$$

and all of the axioms which apply to μ also apply to λ.

*University of Carnegie Mellon, Pittsburgh, PA

Let F be a forgetful functor from $\mathcal{X}$ to $\mathcal{Y}$. Then let $\mathcal{X}$ be a vector space homomorphic to $\mathbb{R}^{d_1} \otimes \mathbb{Q}^{d_2} \otimes \mathbb{Z}^{d_3} \otimes \mathbb{G}^{d_4} \otimes \mathbb{Z}_{k_1} \otimes \cdots \otimes \mathbb{Z}_{k_\ell}$, where $\mathbb{R}$ denotes the real numbers, $\mathbb{Q}$ denotes the rationals, $\mathbb{Z}$ denotes the integers (and $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$ as usual). Let $\mathcal{Y} \subset \mathbb{R}^{d_5}$ be the range of outcomes from $\mathcal{X}$ under some unknown continuous, bounded map $f : \mathcal{X} \to \mathbb{R}$. We are interested in inferring the map f from some random observations $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n) \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}}$, where the distribution $\mathbb{P}$ is unknown. One way to do that is to pass the observed covariates X_i through a *forgetful functor* as is used in category theory.

Definition. Let $\mathcal{C}$ be a category based on sets. We say a functor F is a *forgetful functor* from $\mathcal{C}$ to another category $\mathcal{D}$ if for every morphism $\mu \in \{D \to D\}$ there exists a morphism $\lambda \in \{C \to C\}$ such that

$$F(\lambda) = \mu$$

and all of the axioms which apply to μ also apply to λ.

Let F be a forgetful functor from $\mathcal{X}$ to $\mathcal{Y}$. Then let $\mathcal{X}$ be a vector space homomorphic to $\mathbb{R}^{d_1} \otimes \mathbb{Q}^{d_2} \otimes \mathbb{Z}^{d_3} \otimes \mathbb{G}^{d_4} \otimes \mathbb{Z}_{k_1} \otimes \cdots \otimes \mathbb{Z}_{k_\ell}$, where $\mathbb{R}$ denotes the real numbers, $\mathbb{Q}$ denotes the rationals, $\mathbb{Z}$ denotes the integers (and $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$ as usual). Let $\mathcal{Y} \subset \mathbb{R}^{d_5}$ be the range of outcomes from $\mathcal{X}$ under some unknown continuous, bounded map $f : \mathcal{X} \to \mathbb{R}$. We are interested in inferring the map f from some random observations $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n) \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}}$, where the distribution $\mathbb{P}$ is unknown. One way to do that is to pass the observed covariates X_i through a *forgetful functor* as is used in category theory.

Definition. Let $\mathcal{C}$ be a category based on sets. We say a functor F is a *forgetful functor* from $\mathcal{C}$ to another category $\mathcal{D}$ if for every morphism $\mu \in \{D \to D\}$ there exists a morphism $\lambda \in \{C \to C\}$ such that

$$F(\lambda) = \mu$$

and all of the axioms which apply to μ also apply to λ.

Let F be a forgetful functor from $\mathcal{X}$ to $\mathcal{Y}$. Then let $\mathcal{X}$ be a vector space homomorphic to $\mathbb{R}^{d_1} \otimes \mathbb{Q}^{d_2} \otimes \mathbb{Z}^{d_3} \otimes \mathbb{G}^{d_4} \otimes \mathbb{Z}_{k_1} \otimes \cdots \otimes \mathbb{Z}_{k_\ell}$, where $\mathbb{R}$ denotes the real numbers, $\mathbb{Q}$ denotes the rationals, $\mathbb{Z}$ denotes the integers (and $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$ as usual). Let $\mathcal{Y} \subset \mathbb{R}^{d_5}$ be the range of outcomes from $\mathcal{X}$ under some unknown continuous, bounded map $f : \mathcal{X} \to \mathbb{R}$. We are interested in inferring the map f from some random observations $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n) \in \mathbb{P}_{\mathcal{X} \times \mathcal{Y}}$, where the distribution $\mathbb{P}$ is unknown. One way to do that is to pass the observed covariates X_i through a *forgetful functor* as is used in category theory.

Definition. Let $\mathcal{C}$ be a category based on sets. We say a functor F is a *forgetful functor* from $\mathcal{C}$ to another category $\mathcal{D}$ if for every morphism $\mu \in \{D \to D\}$ there exists a morphism $\lambda \in \{C \to C\}$ such that

$$F(\lambda) = \mu$$

and all of the axioms which apply to μ also apply to λ.

We can now begin to implement our method.

2 Method

The target of our estimation is the continuous, bounded map $f : \mathcal{X} \to \mathbb{R}$, where for simplicity we suppose $\mathcal{X}$ is homeomorphic to $\mathbb{R}^d$. In the general case the collection of candidate functions is uncountable and unfeasible to search, so instead we limit our fitting procedure to some class $\mathcal{F}$ of, say, L-Lipschitz functions although this choice will not affect convergence rates. Consider the objective function the training error and set

$$\hat{f} = \underset{\tilde{f} \in \mathcal{F}}{\operatorname{argmin}} \left\{ \hat{\mathbb{E}} \left\| \tilde{f}(X) - f(X) \right\|^2 \right\}, \tag{1}$$

where $\hat{\mathbb{E}}[\cdot]$ is with respect to the empirical distribution of the sample X_j. Note that we do not know a priori whether there exists an interpolating $\tilde{f}$ in $\mathcal{F}$, so the training error for $\hat{f}$ is not necessarily zero.

Now define the functor $F : \{X \to \mathbb{R}\} \to \{X \to \mathbb{R}\}$ which takes as its input any classifier $\hat{f}$ and returns the constantly zero classifier z. That is, $x \overset{z}{\mapsto} 0$ for all $x \in \mathcal{X}$. A natural question arises for the astute reader: For which choices of $\hat{f}$ does $F(\hat{f})$ give meaningful predictions? In numerical simulations, the mean squared error for $F(\hat{f})$ is abysmal on most data sets. It turns out that this is due to our inadequate choice of risk measure. Indeed, choosing the appropriate category-theoretical metric for performance, we obtain much more reasonable results, as summarized in Figure 1. We choose instead to evaluate our wrapped classifier according to the risk

$$R(\hat{f}, f) = \mathbb{E} \left\| F(\hat{f})(X) - F(f)(X) \right\|^2. \tag{2}$$

This risk reveals the convenience of our choice of forgetful functor as is used in category theory.

Definition. Let $\mathcal{C}$ be a category based on sets. We say a functor F is a *forgetful functor* from $\mathcal{C}$ to another category $\mathcal{D}$ if for every morphism $\mu \in \{D \to D\}$ there exists a morphism $\lambda \in \{C \to C\}$ such that

$$F(\lambda) = \mu$$

and all of the axioms which apply to μ also apply to λ.

We can now begin to summarize our results.

3 Results

We summarize our results in a physical note to ourselves which we promptly misplace on our way to the printing press.

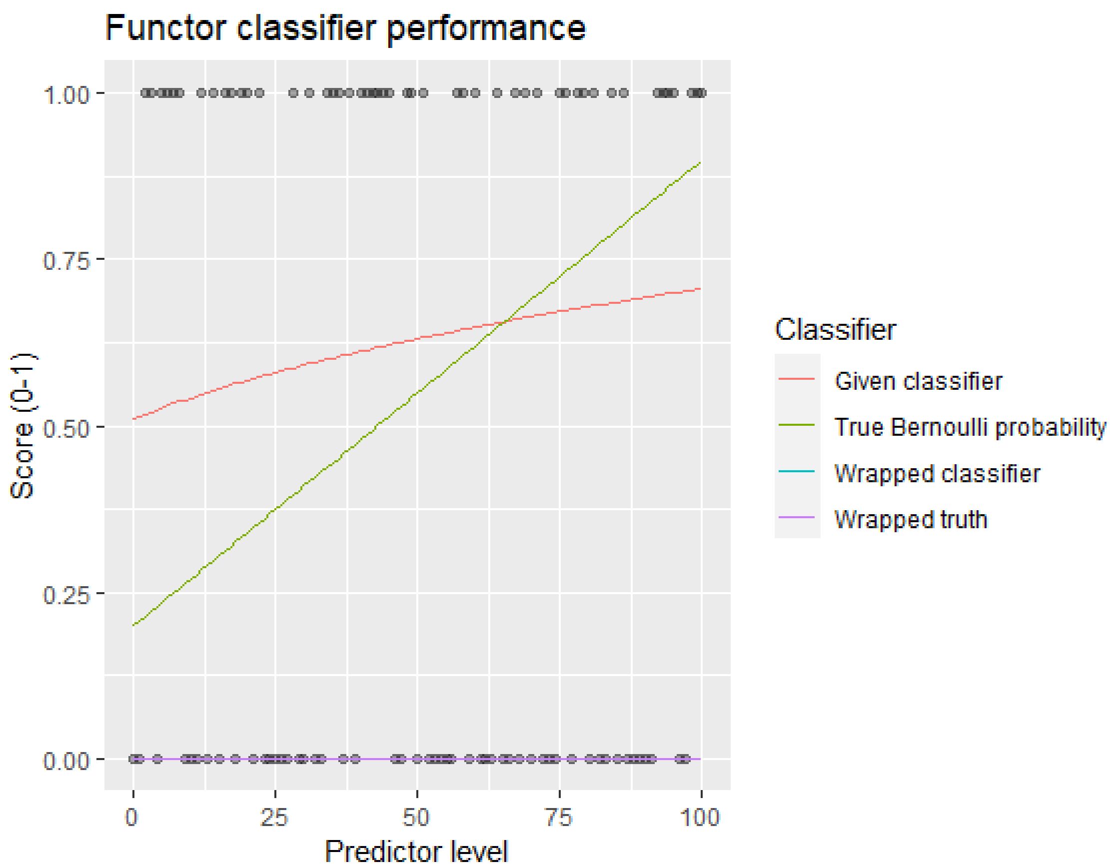

Figure 1: The green curve represents the "truth" $p(Y = 1|X = x)$. The red curve was learned from a sample of one hundred training points. The blue curve is the classifier after wrapping it in the functor F, and the purple curve is the true curve wrapped in F. Although on the training data the wrapped classifier does not perform particularly well, we see that under the view of the functorial wrapper, it identically mimics the truth. This convenient fact has held regardless of how unruly the truth has been specified in numerous simulations.