Primal Formulation of a Nonlinear Classifier With MMSE

Keith H. Bova

***Abstract*—This paper examines the effectiveness of a non-linear classifier that separates a binary dataset by applying a Volterra expansion–projecting it into $\mathbb{R}^{10}$. This classifier creates a boundary that can classify almost all points.**

***Index Terms*—Volterra Expansion, Non-Linear Classifier**

I. Introduction

The Volterra series of a non-linear function is a transformation that maps two functional spaces. Similar to the Taylor series, the Volterra series is expressed as an infinite sum; however, unlike the Taylor series, it is based on convolutional integrals.

A. The Volterra Expansion

A volterra series can be expressed as:

$$\hat{y} = \sum_{k=0}^{K} a_k x^k \tag{1}$$

If we consider a space of $\mathbb{R}^2$, and we want to pass from this space into $\mathbb{R}^p$ such that:

$$p = \binom{2+3}{3} = 10 \tag{2}$$

We first formulate the expansion:

$$\langle 1, x_1, x_2, x_1^2, x_2^2, x_1x_2, x_1^2x_2, x_1x_2^2, x_1^3, x_2^3 \rangle \tag{3}$$

Which becomes:

$$\underline{\mathbf{x}} = \begin{bmatrix} 1 \\ x[n] \\ x[n-1] \\ x^2[n] \\ x^2[n-1] \\ x[n]x[n-1] \\ x^3[n] \\ x^3[n-1] \\ x^2[n]x[n-1] \\ x[n]x^2[n-1] \end{bmatrix} \tag{4}$$

II. Experiments

A. Data Generation

For this experiment, we generate data using a Gaussian distribution and a parameter α that causes convolutional spread. The lower the value of α, the closer classes are to each other.

B. Classifying Low Values of α

For this experiment, we first begin by plotting our data in $\mathbb{R}^2$. We generate N data samples with a convolutional spread constant α, such that:

$$\alpha = 0.2, N = 100 \tag{5}$$

The two classes are shown below.

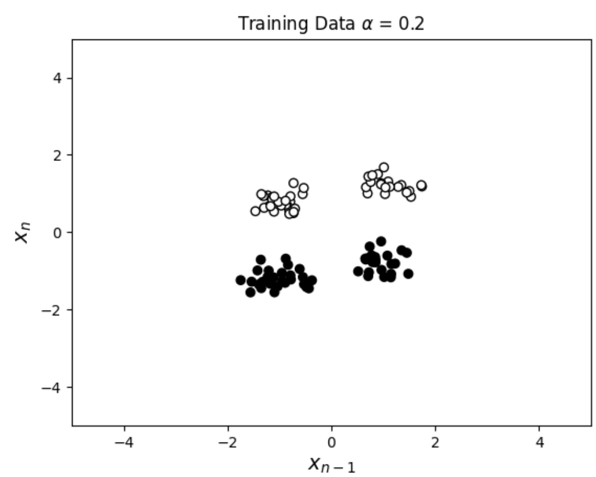

Fig. 1. Data clusters in $\mathbb{R}^2$ using $\alpha = 0.2, N = 100$

We can visually see that the two classes are linearly separable, and a support vector machine, support vector classifier would work well in this case; nonetheless, we apply the non-linear classifier and can see that the non-linear approach works in this case also.

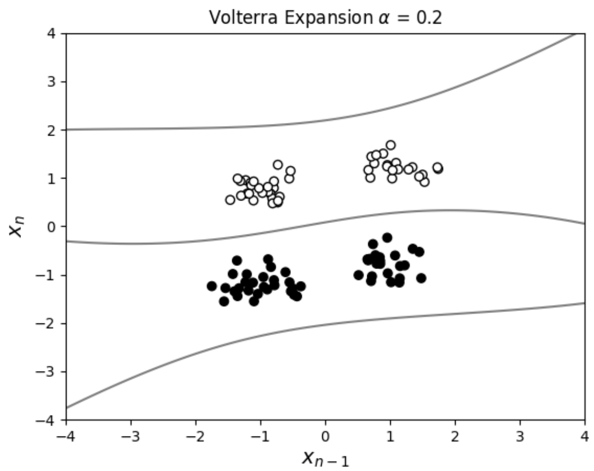

Fig. 2. Example in $\mathbb{R}^2$ classifying linearly separable data with a non-linear classifier

C. Classifying With High Values of α

For this experiment, we plot our data in $\mathbb{R}^2$ using a new convolutional spread constant α:

$$\alpha = 1.5, N = 100 \tag{6}$$

We can see in the figure below that our data is no longer linearly separable.

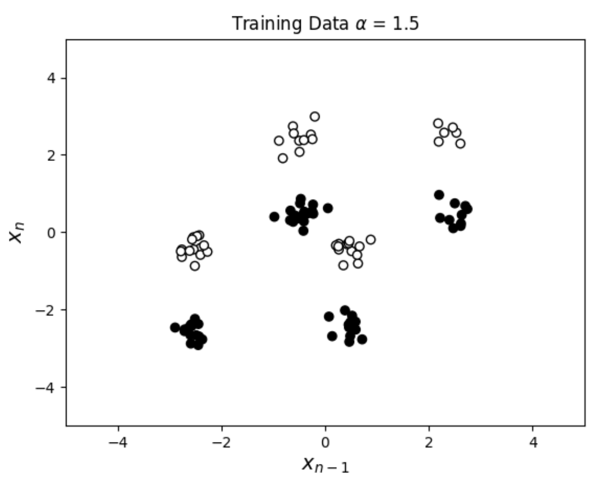

Fig. 3. Data clusters in $\mathbb{R}^2$ using $\alpha = 1.5, N = 100$

We now pass our data through the non-linear classifier, and draw the contour.

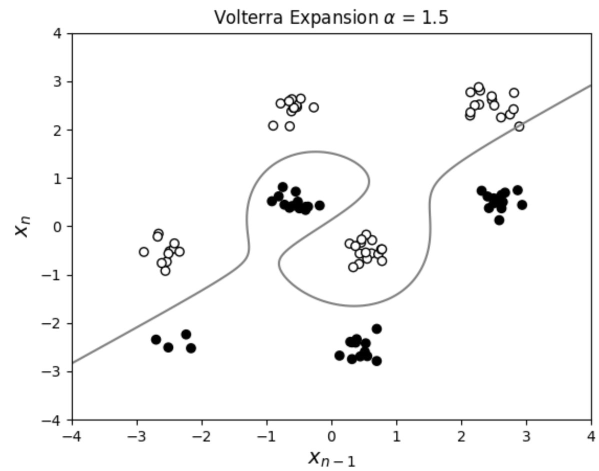

Fig. 4. Applying the non-linear classifier to our data $\alpha = 1.5, N = 100$

III. Conclusion

We see from the following experiments that the non-linear classifier works separating data were the linear classifier does not; however, it is important to note that underfitting does occur in this case.

IV. Source Code

https://github.com/keithhbova/support_vector_machines/