Non-Linear Support Vector Machine Classifiers

Keith H. Bova

***Abstract*—This paper examines three kernel dot products for classifying data in $\mathbb{R}^2$: the linear kernel, the third order kernel, and the radial basis (square exponential) kernel. Source code to reproduce the experiments is given at the end of the paper.**

***Index Terms*—Kernel technique, regression, vector classifier, linear kernel, third-order kernel, radial basis function**

I. Introduction

A. Linear Kernel

We represent the kernel as:

$$k(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Z}}}) = \underline{\underline{\mathbf{X}}}^T \underline{\underline{\mathbf{Z}}} \tag{1}$$

B. Third Order Polynomial

We represent the kernel as:

$$k(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Z}}}) = ||\underline{\underline{\mathbf{X}}} - \underline{\underline{\mathbf{Z}}}||^3 \tag{2}$$

C. Radial Basis

We represent the kernel as:

$$k(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Z}}}) = e^{\frac{-||\underline{\underline{\mathbf{X}}} - \underline{\underline{\mathbf{Z}}}||^2}{2\sigma^2}} \tag{3}$$

II. Experiments

A. Linear Kernel

Here, we apply the linear classifier to the data that is linearly separable and we see a result that is exactly what we expect.

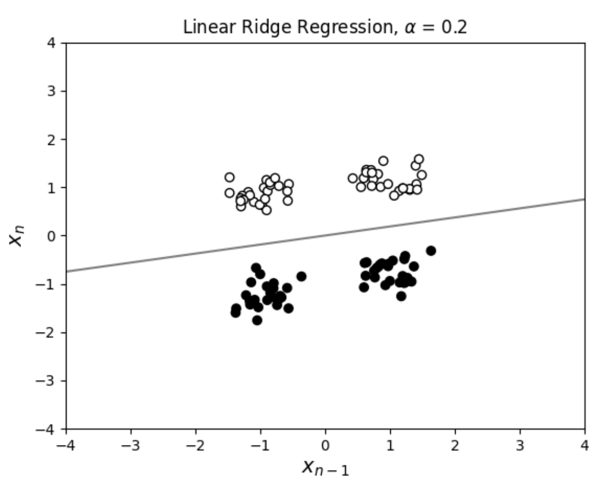

Fig. 1. Example in $\mathbb{R}^2$ with a linear classifier separating the two classes

While this model works well for the lower values of α, it breaks down when α increases and the data is no longer linearly separable with zero errors. The kernel is not usable in this case.

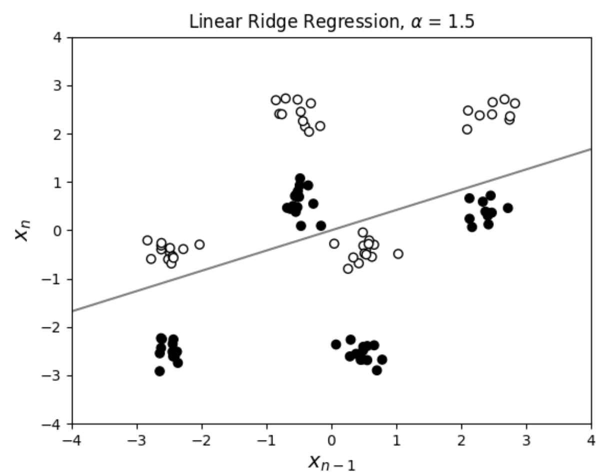

Fig. 2. Example in $\mathbb{R}^2$ where the data is not linearly separable with zero errors

B. Nonlinear Third-Order Polynomial

Because the linear kernel does not accurately classify the data with a larger value of α, we must find a new solution. Here, we apply the nonlinear kernel–and we see that with the first distribution, we still achieve a result similar to the linear classifier.

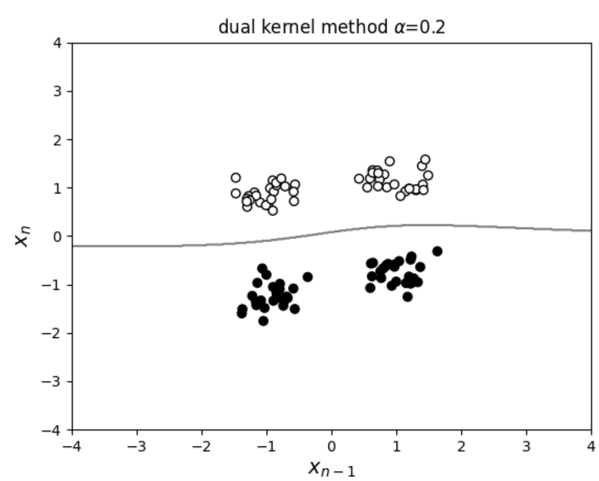

Fig. 3. Applying the nonlinear kernel to our data $\alpha = 0.2, N = 100$

Unlike the linear classifier–however–the third order polynomial can accurately separate the distribution with the larger value for α.

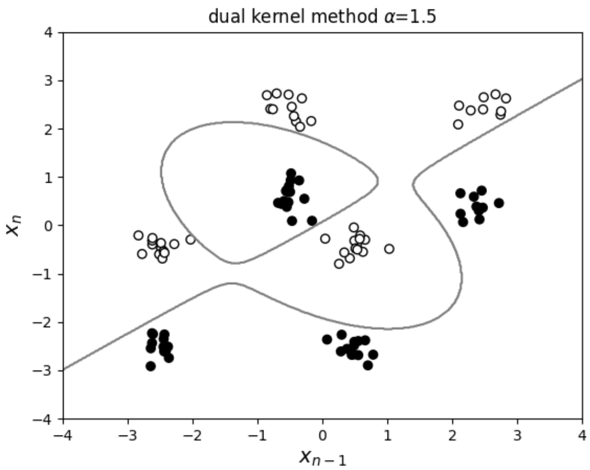

Fig. 4. Applying the nonlinear kernel to our data $\alpha = 1.5, N = 100$

C. Radial Basis Kernel

The third order polynomial can capture the distribution of our data with a reasonable degree of accuracy; however, this kernel underfits. When we apply the radial basis kernel on our first distribution with the lower value of α we see the following:

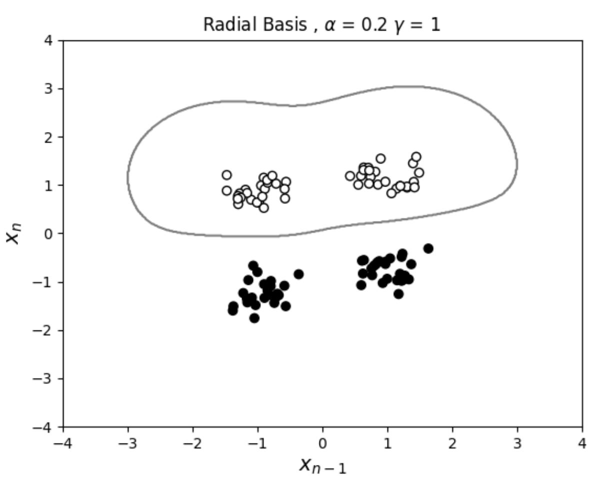

Fig. 5. Applying the radial basis kernel to our data $\alpha = 0.2, N = 100$

At first glance, this result seems less promising; however, when applied to the data that is not linearly separable with zero errors, we achieve a contour that captures the data with a higher degree of accuracy–and does not appear to underfit our distribution.

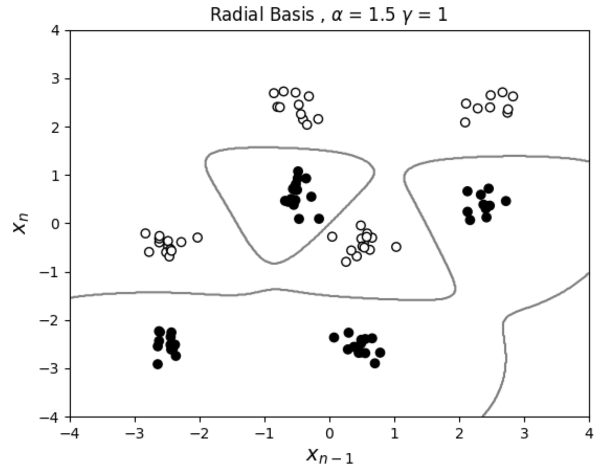

Fig. 6. Applying the radial basis kernel to our data $\alpha = 1.5, N = 100$

III. Conclusion

The radial basis kernel and the polynomial kernel both can accurately capture a nonlinear distribution; however, in the case of our data, the radial basis kernel appears more accurate. This is likely because the original distribution in our data function follows a gaussian–which is being approximated with the kernel we are using to generate the contour.

IV. Source Code

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