1. Objective of whole project.

Compare MADGE data to other machine learning algorithms like SVM and neural networks. Accuracy and speed will be the two most optimized features, with accuracy without overfitting as the most interesting piece to explore.

1. What was implemented last week.

A comparison of various sigmas was done with two methods.

- The first method used a normal distribution calculation:

(s/(s+t))\*Normal(x, W/ σ) + t/(s+t))\*Normal(x, D/ σ)

- The second method uses an absolute factor, with σ as a dividing factor.

(s/(s+t))\*W/σ) + t/(s+t))\*D/σ)

- Spiral Data was stretched by various factors, both in the X and Y direction. Data is labeled with “100-10” for example stretches the X data by a factor of 100 and Y data by a factor of 10.

Using the first method, the following data was constructed:

It seems as though for small stretching, there is an optimal sigma that is valid here. But for larger stretch factors, there is no optimal sigma that would generate an accuracy of > 60 %.

Using the second method, the following graphs were generated:

I used three graphs since I wanted to use a large range of sigmas to test optimization.

Figure 1.

From Figure 1, it seems as though a larger sigma is required for a higher stretched data set.

Figure 2.

Figure 2 seems to indicate that there maximum sigma does have a drop in accuracy at a certain point around the “stretch-factor”.

Figure 3.

I made an assumption that a large sigma would be required for a large stretch factor, but it seems as though a sigma of 100 roughly had the optimal accuracy, even at higher stretch factors.

1. Plans for upcoming week.

Trying to figure out if there is a range on sigmas that need to be tested. At the worse, this is an empirical binary search, at best it’s an approximation we can use an equation to determine.

1. Objective of whole project.

Compare MADGE data to other machine learning algorithms like SVM and neural networks. Accuracy and speed will be the two most optimized features, with accuracy without overfitting as the most interesting piece to explore.

1. What was implemented last week.

Attempted to use a best calculation for sigma.

- Tensorflow data is manipulated to be stretched out to attempt to test sigmas

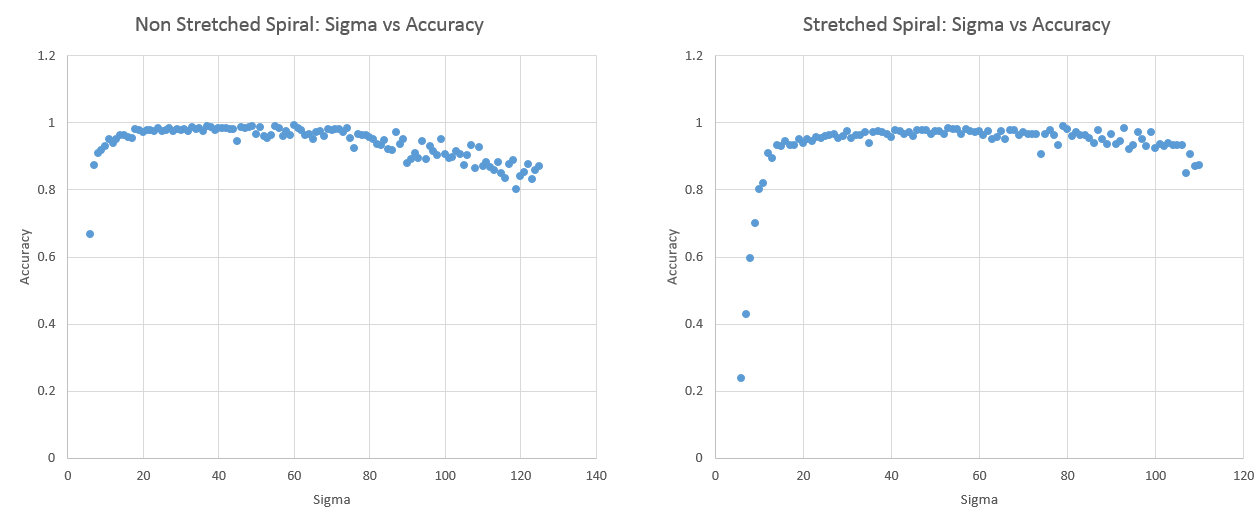
- Used the equation for calculating sigma between two points with distance calculation based on dimensions:

Where d is the distance from the point on dimension j to point M, D is the range of dimension j, and the sum of n is the total sum of all the distances for each dimension.

Sigmas were plotted and calculated for a given range.

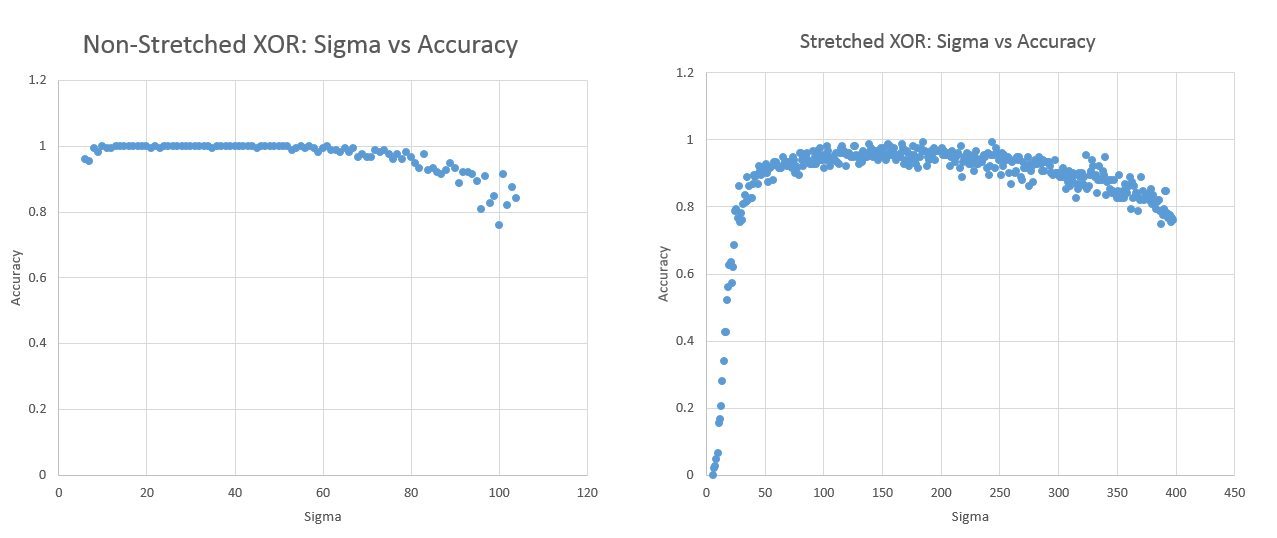
The spiral data was stretched out and an optimal sigma was found with testing.

Given the right sigma, accuracy is able to be in the 90s.



Two Spiral Stretched: 831.3599

Two Spiral Non-Stretch Range: 8.247



XOR Stretch Range: 3178.55297

XOR Non-Stretch Range: 8.2470265

Stretch range is the sum of ranges from all dimensions.

1. Plans for upcoming week.

I think I need to test the stretching range of the data for one data set to determine if there is a correlation.

Different data sets at different stretch ranges should have similar sigmas?

Apply another method that encapsulates the “shape” of the data better.