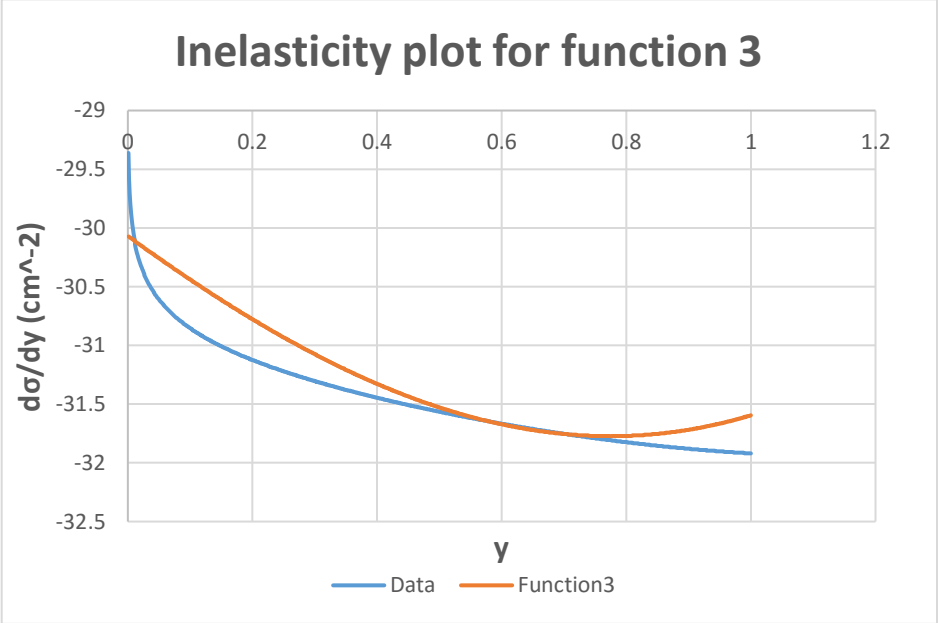
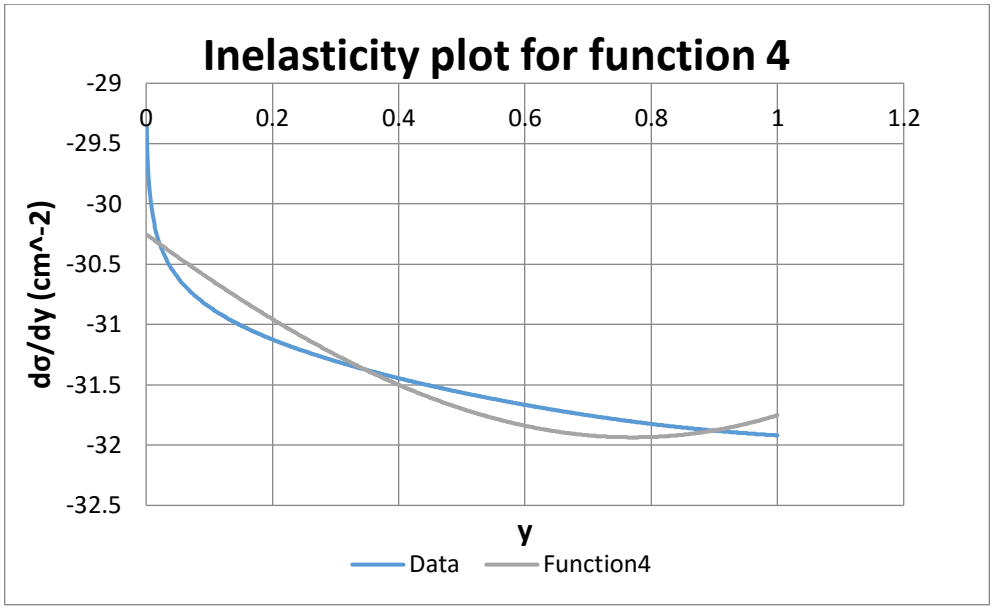
Inelasticity data analysis using karoo gp

This is a document describing my project with karoo machine learning and using it to describe the inelasticity distribution from the Connolly et al. cross-section paper. At first, karoo was performing poorly on the data set but then later, I modified the operators file to include trig and exp functions and also included too many coefficients. It was only after decreasing the number of coefficients I was using and modifying the regression parameters that I started to get reasonable functions. Karoo basically outputted 4 different functions which had the best fit. These four functions are labeled as "Function 3", "Function 4", "Function 5" and "Function 6" on my graphs. I plotted them against the inelasticity data for 10^{12} energy.

Function 3 is: (0.3496*y) + (0.8306*EXP(y)*SIN(y)) - (5.0335*SIN(y)) - (COS(y)) - (29.0687). As you can see the function fits the data really well between y values of 0.5 and 0.7 but it still needs a lot of improvement for lower values of y.

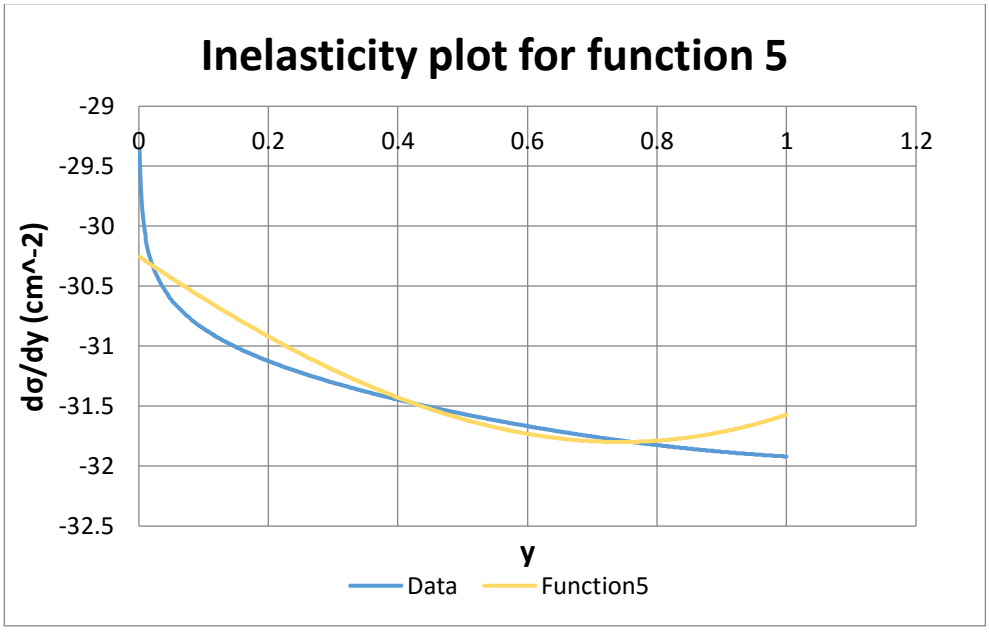


Function 4 is: (0.3756*y) + (0.8306*EXP(y)*SIN(y)) - (5.0335*SIN(y)) - (COS(y)) - (29.252).

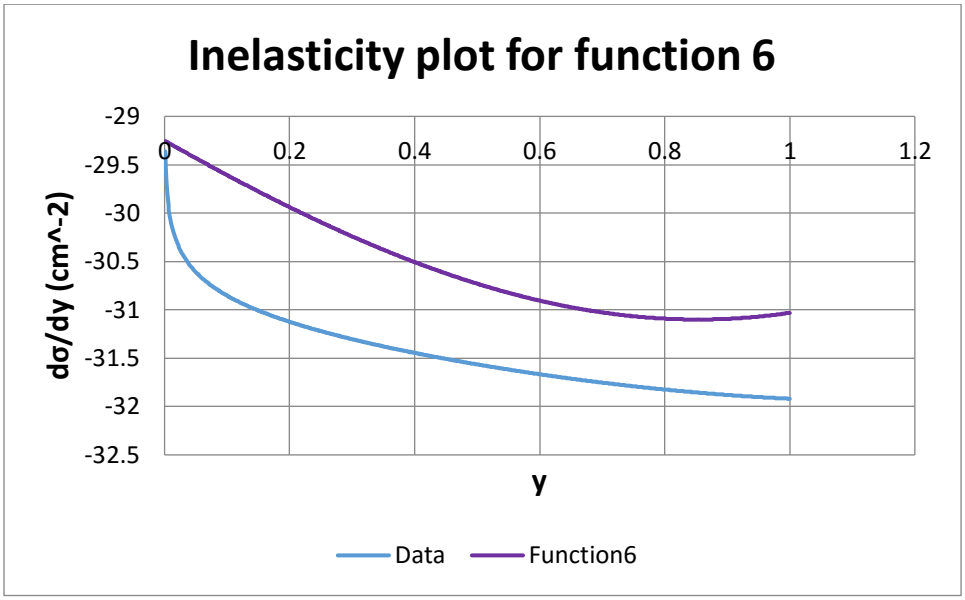


Function 4 and 5 had the least RMS error when I calculated it to see what the error between the data and function was. The error was like 0.14 and 0.15 for each. For the rest, it was a little higher. For function 3 it was 0.22 and function 6 it was 0.94.

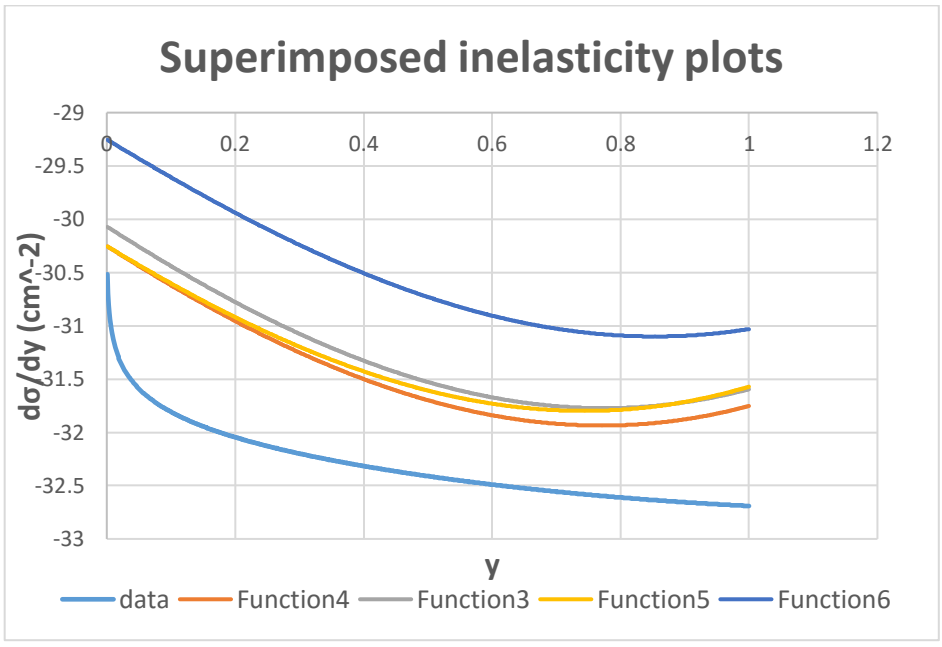
Function 5 is:(0.5569*y) + (0.8306*EXP(y)*SIN(y)) - (5.0335*SIN(y)) - (COS(y)) - (29.252)



Function 6 is: (0.5569*y) + (0.8306*EXP(y)*SIN(y)) - (5.0335*SIN(y)) - (29.252). So the only difference in this graph from the previous one is that I removed the cos term as the shape seemed to improve but then it still doesn't fit the data. I think varying the parameters might help but I haven't tried it yet so I don't know for sure if that'll work.



The superimposed inelasticity plots include the same four functions as above but I used an energy of $10^{8.6667}$ for my data. As you can see, they fit the data at none of the y values.



This is what one would expect as energy is also a variable in this case and could be fed into karoo as well to give functions based on which energy we're looking at.

In the future, I think it would be helpful to divide the data into low y values and high y values and try running karoo because currently it's not doing a good job on low values of y for some reason. Also, if we could somehow make the function steeper, that would also be great. I'm also working on improving karoo's fitness functions so that the most fit functions will look more like our data. That is it for now and I will keep all of you posted regarding anything new.