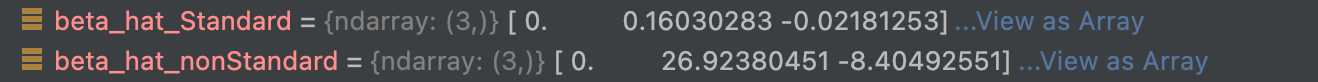
Please see code in the .py files.

Chart, line chart, scatter chart

Description automatically generated

If the data are separable and the two clusters are far enough, the two algorithms both work but will give different lines.

If we standardize the beta, the gradient will be zero if the norm of beta is large and the gradient will be extremely large if the norm is small enough. If the norm is close to zero the gradient will be NaN. The algorithm will no longer work.



Derivative calculation:

A piece of paper with writing on it

Description automatically generated with medium confidence