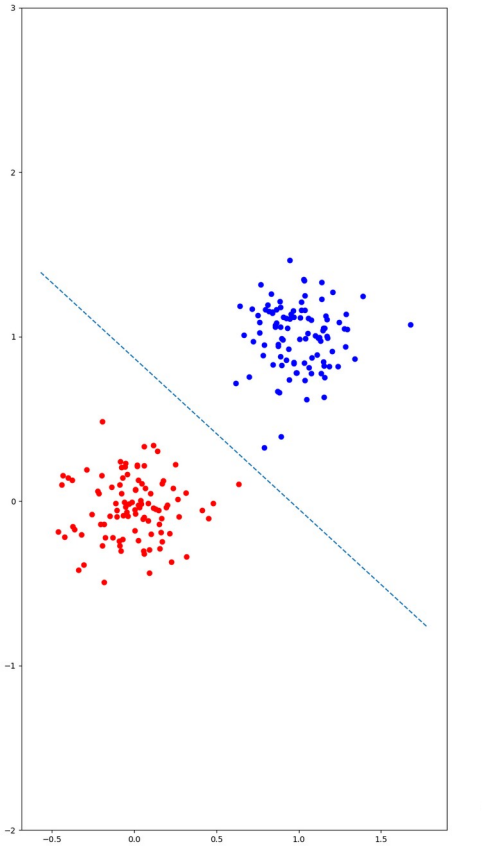
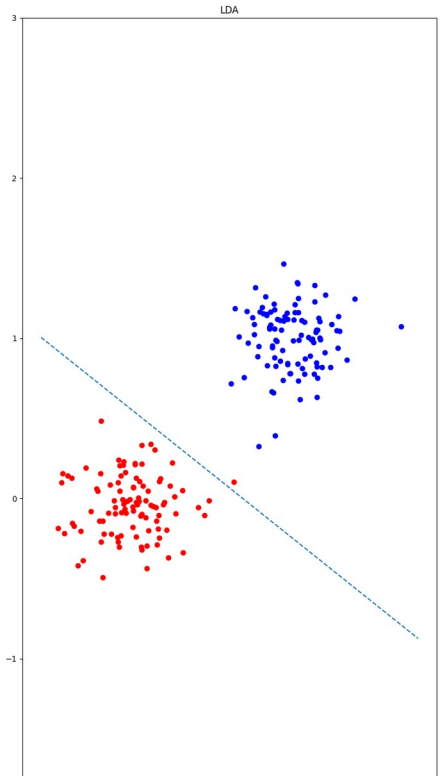
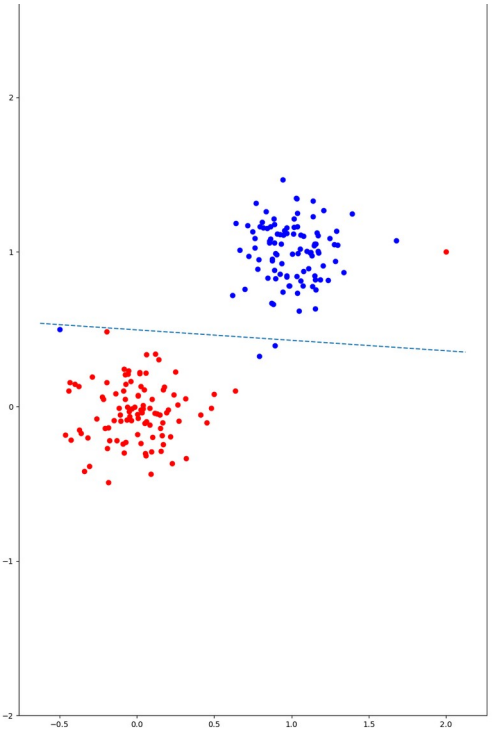
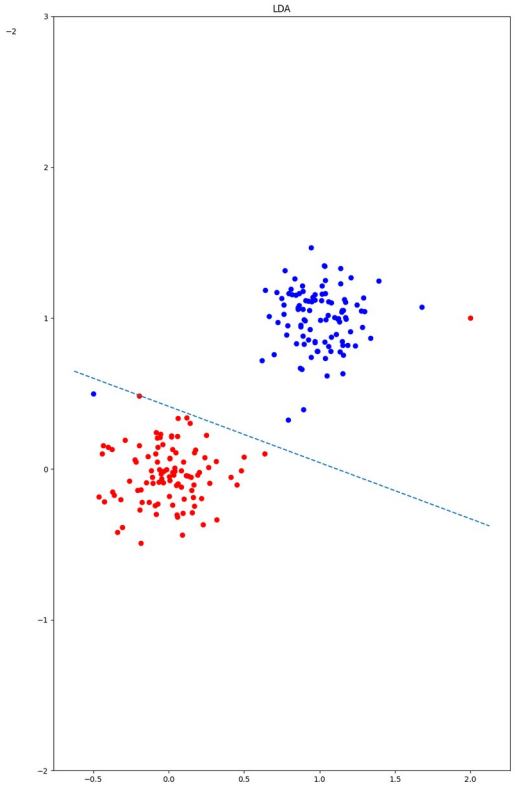
Devon Callanan
Homework 2
Pattern Recognition

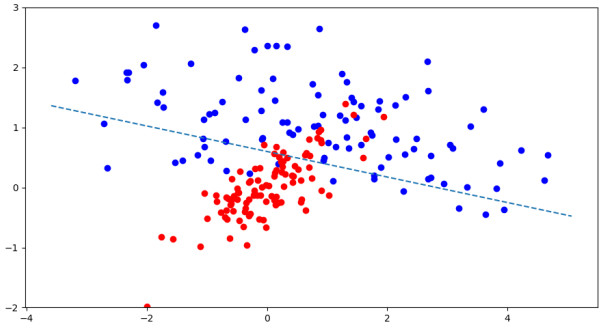
1. Logistic regression on the left, LDA on the right

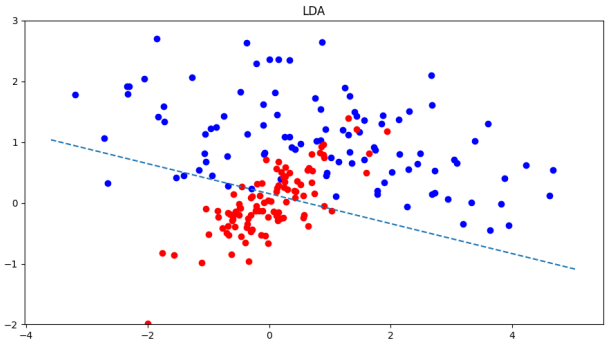


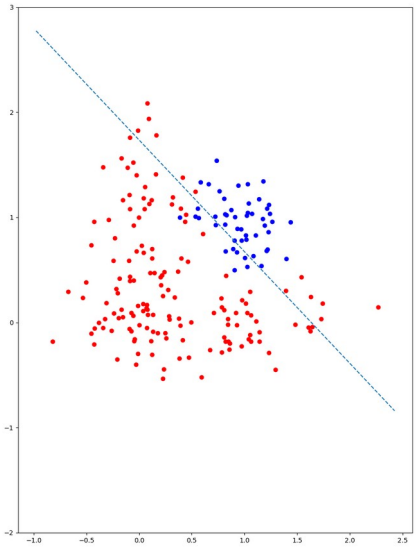


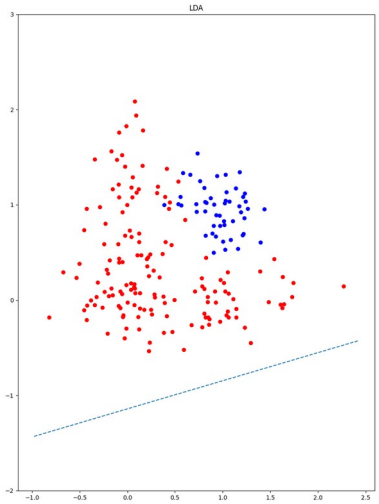








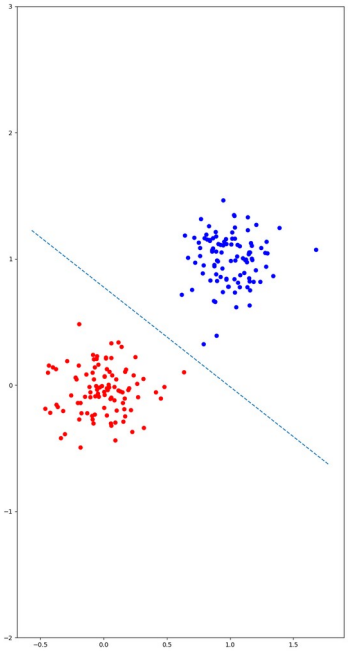


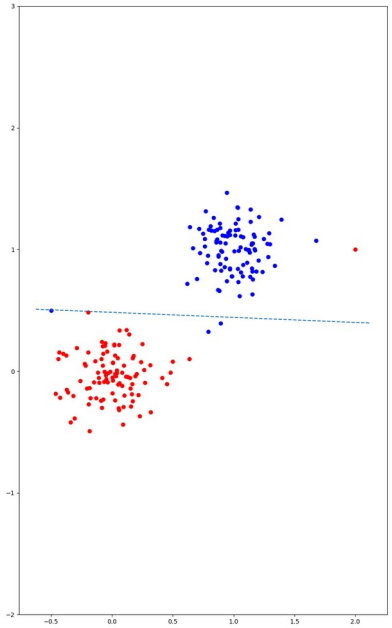


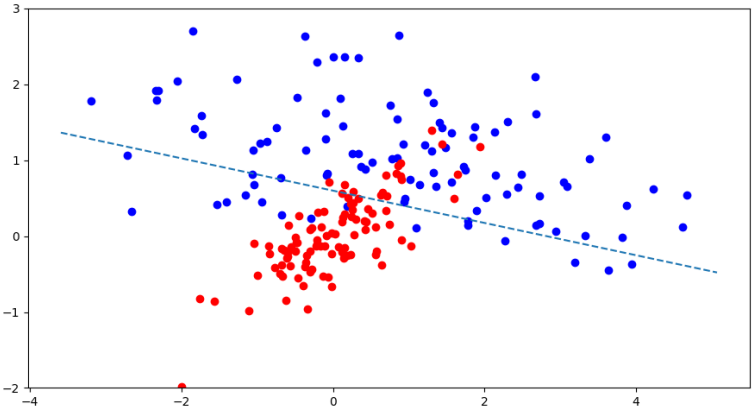
c) A high alpha could lead to oscillation of the gradient descent which would never converge and a low alpha would slow convergence. Tests bellow were

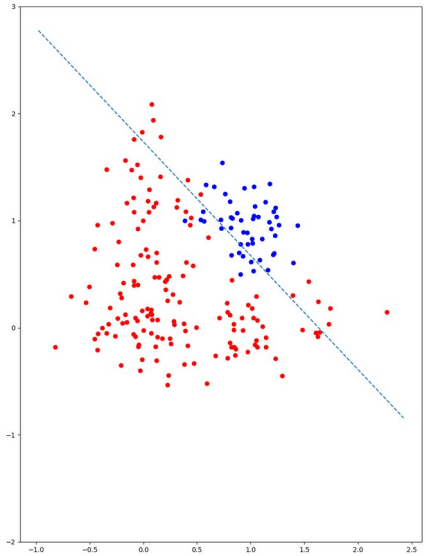
dataset	1	2	3	4
iterations	42	36	20	49
Accuracy	1	.99	.82	.88

2.



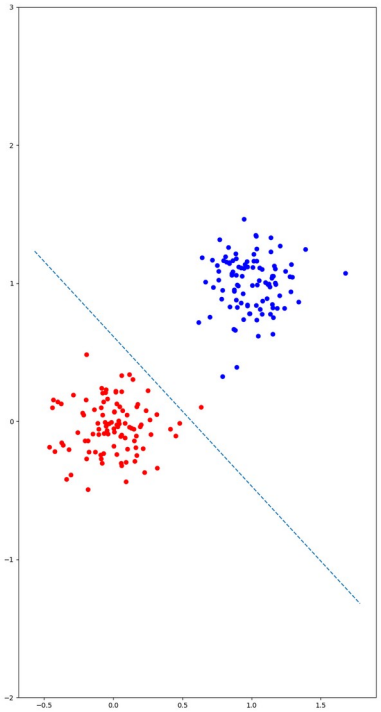


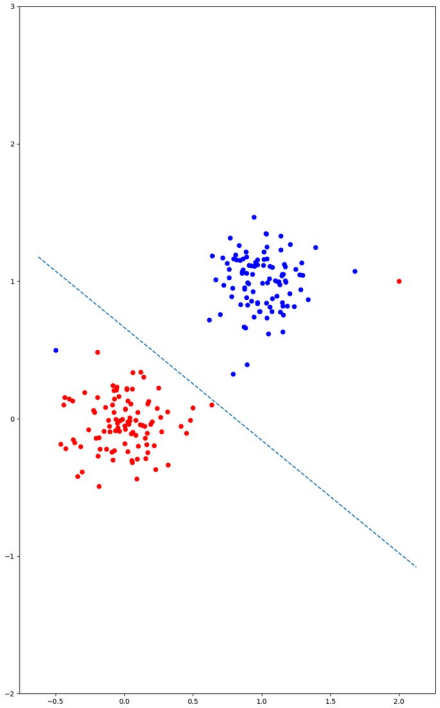


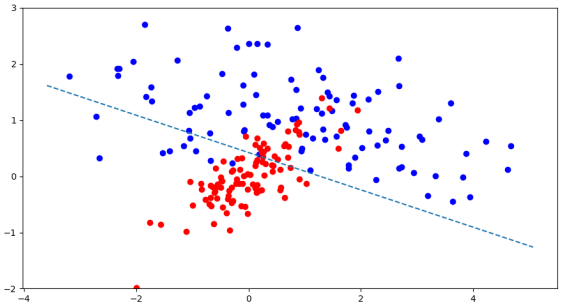


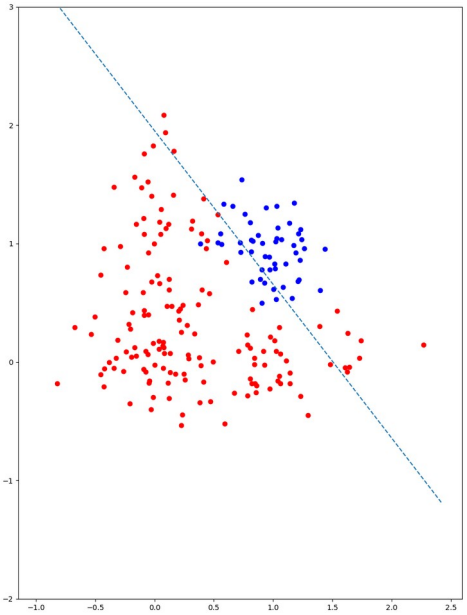
dataset	1	2	3	4
iterations	8	5	4	5
Accuracy	1	.98	.825	.885

3.









dataset	1	2	3	4
iterations	~120	~130	~200	~500
Accuracy	.995	.99	.825	.86

Discussion

Bellow is a plot showing the log likelyhood as the solutions reach convergence. For all other data, iterations were stopped with the change in the log likelyhood, denoted as ‘delta’ was bellow a threshold. This was done to signify that the progress toward the minimum had slowed.

While the regular logistic regression and newtonian tuned regression converged quickly and consistently, the stochastic descent was much less predictable. Depending on the order of points used in gradient calculation, the convergence could arrive quickly or slowly to either an ideal or non-ideal solution. While the iterations count is much larger, each iteration is only a single point creating a gradient, not the entire set. With this in mind a convergence is reached with only a couple of loops through the entire dataset.