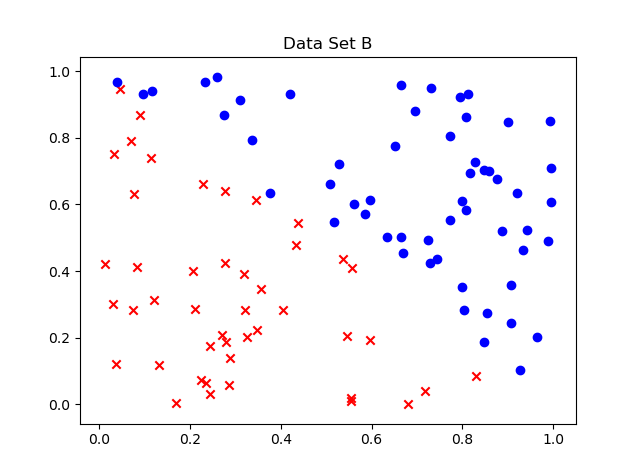
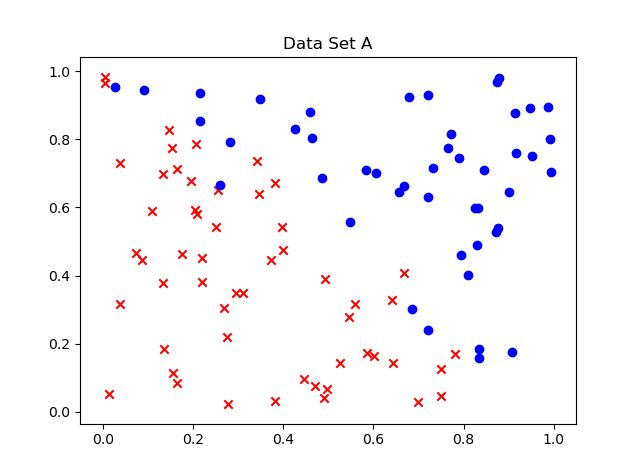
PS2

1a. data set A’s theta converges to a single value after 30372 iterations while B does not.

1b.



Case B is more linearly separable than is Case A.

In the case of linearly inseparable examples, because some points do not follow the linear boundary, the algorithm essentially tries to maximize the number of correctly placed points on either side of the boundary (by maximizing the likelihood of certain events occurring given some parameters). That’s why logistic regression works so well, since theta is bound by non-linear points.

However, for linearly separable data, as in Case B, because of the degree of freedom of theta, the confidence of the functional margin can be infinite without essentially having much effect on the decision boundary: can be maximized with while changes very minutely at high values of , because the sigmoid function converges to 0 or 1. This makes it so that theta never converges.

1c.

i) No. Changing the learning rate is only affects the steps delta theta takes every iteration

ii) Yes. This will cause theta to change very minutely at high iterations, and allowing it to converge (prev\_theta – theta < 1e-15)

iii) No. This will not change the linearly separable trait of Case B, so theta will never converge

iv) Yes. This will allow theta to not just take multiples of itself to change its confidence.

v) Yes. This will make the data less linearly separable.

1d.

No, SVMs use the hinge-loss function that sets the output to 0 if the confidence goes below 1. It also safeguards scaling of the weight and bias by adding the penalty term.