

Logistic Regression: Training stability

- (a) Training model on dataset A costs far more less time than that on dataset B, which means that training on dataset B doesn't converge.
- (b) Let's plot the training results after 10000, 20000, 30000, 40000 iterations.

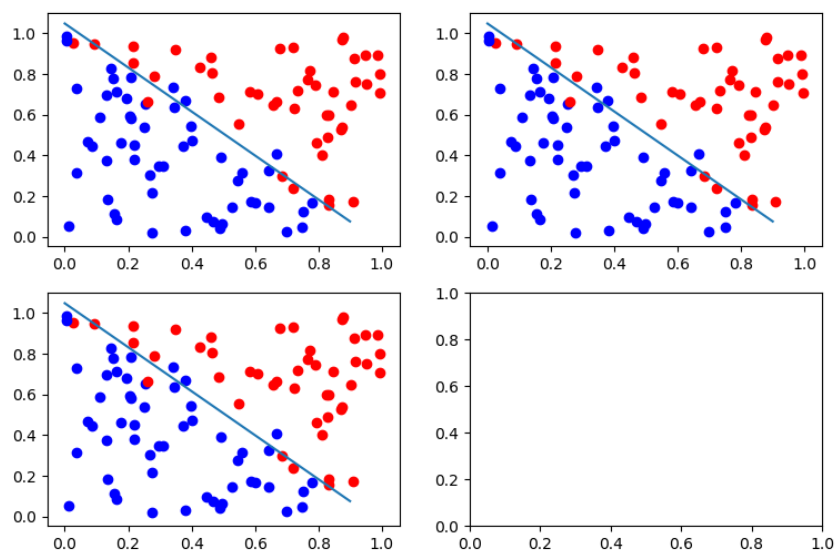


Figure 1: Training Results on Dataset A

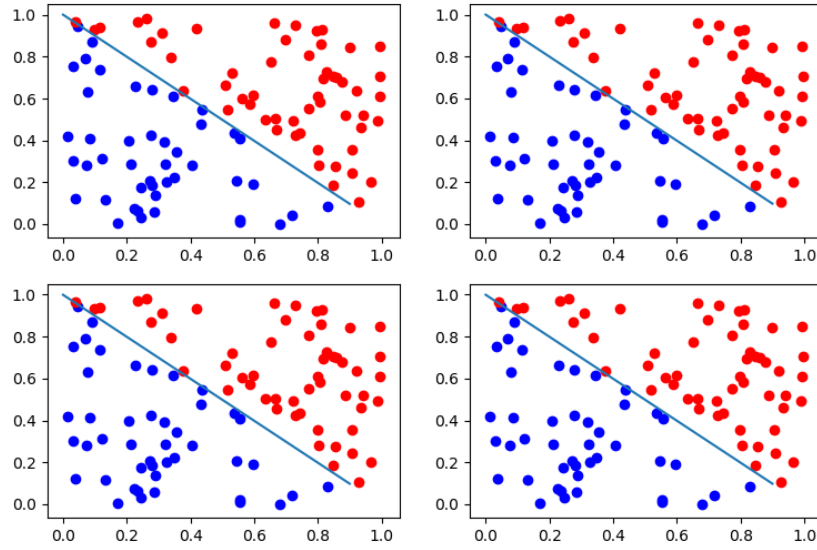


Figure 2: Training Results on Dataset B

From the above two figures, we can see that data on dataset B is hardly to separate (Bad Linearly Separability), which may be the main issue resulting nonconvergence.

- (c)
 - i No. Using a different learning rate only changes the learning speed here, but it won't change the fact that the algorithm has to find the hyperline in hardly separable data.
 - ii No. The same to the former.
 - iii Yes. It will stop $||\theta||$ being infinitely large.
 - iv No. It doesn't change the linearly separability.
 - v Yes. It will expand the feature space, which may let the data linearly separable.
- (d) Yes. It will put the data into a new feature space, where the data may become linearly separable.

Model Calibration

(a) gg