

Case 1:

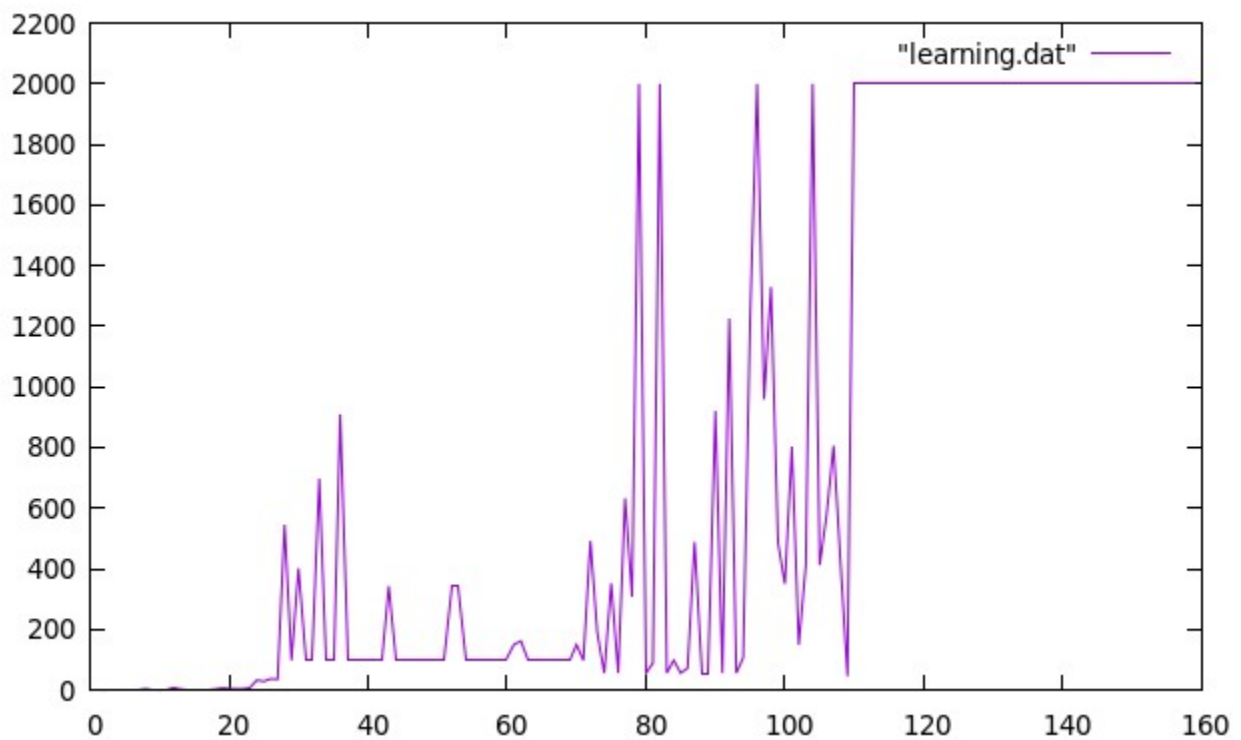
alpha 0.1

gamma 0.95

beta 0.2

LAMDA 0.9(learning1.dat)

Due to lesser learning rate the model takes too much time to learn, The graph converges after fewer iterations but there is an issue with time complexity.

**Case 2:**

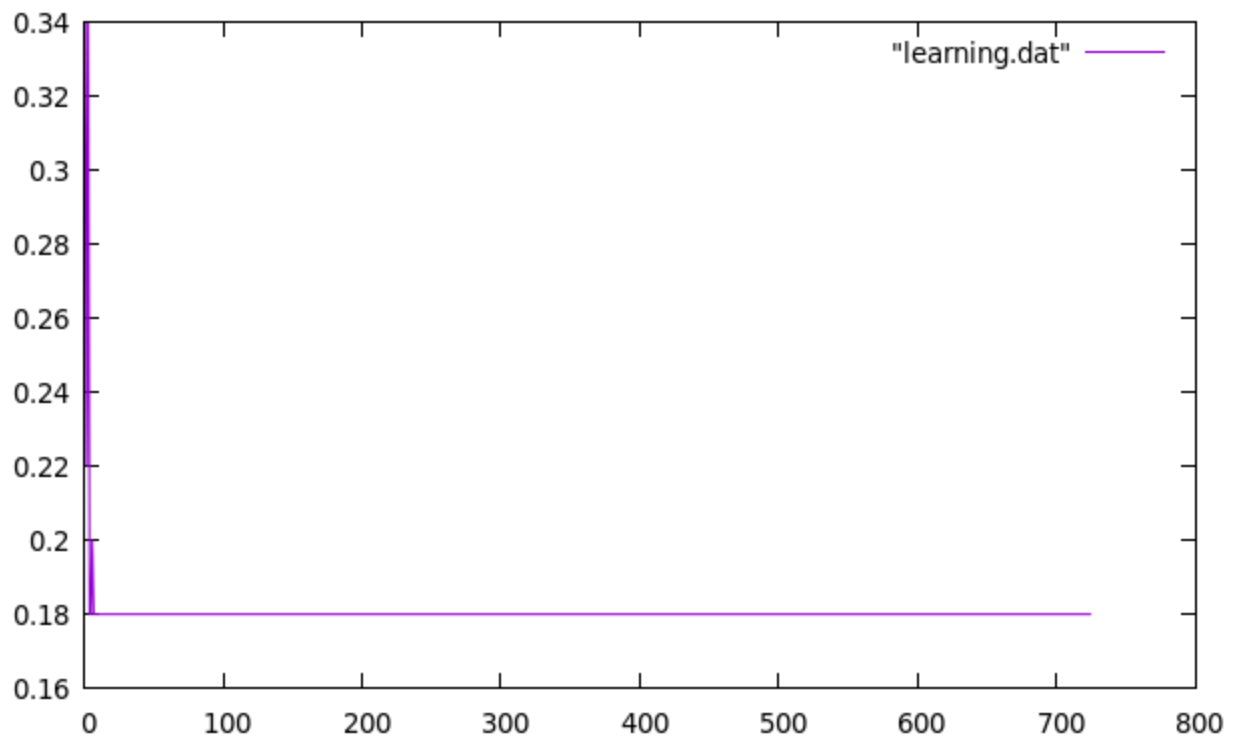
alpha 0.9

gamma 0.95

beta 0.9

LAMDA 0.9(learning2.dat)

By increasing alpha and beta the graph converges and performs better



Case 3:

alpha 0.1

gamma 0.95

beta 0.9

LAMDA 0.9(learning3.dat)

When all the parameter values are similar and only learning rate has been decreased again it took so much time to converge

