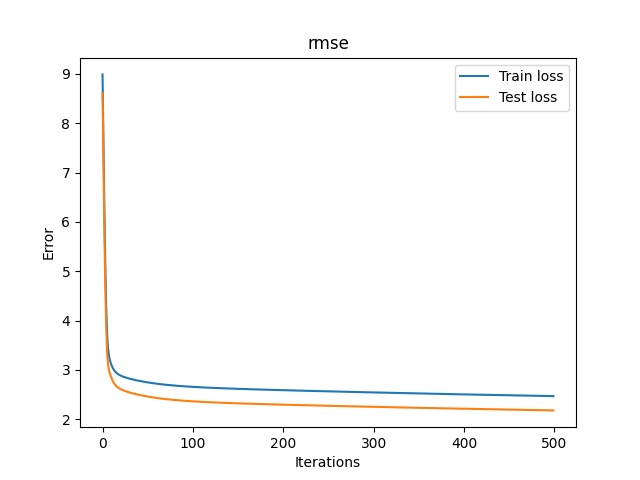
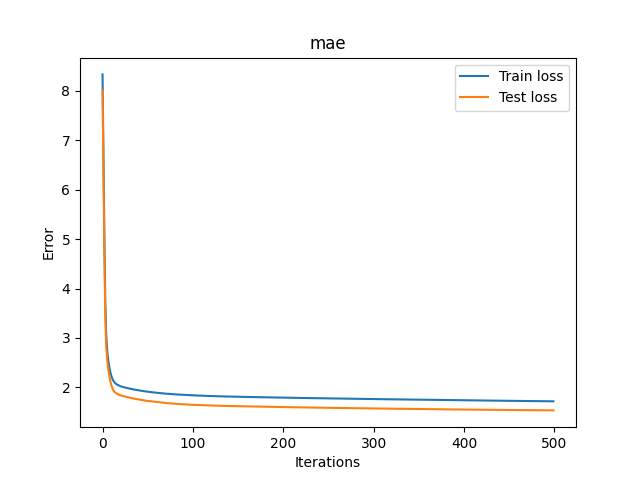
1. Here I have chosen K=10 comparing all the avg testing error on all the values of the K from 2 to 10. Screenshots of the same have been put in the folder.

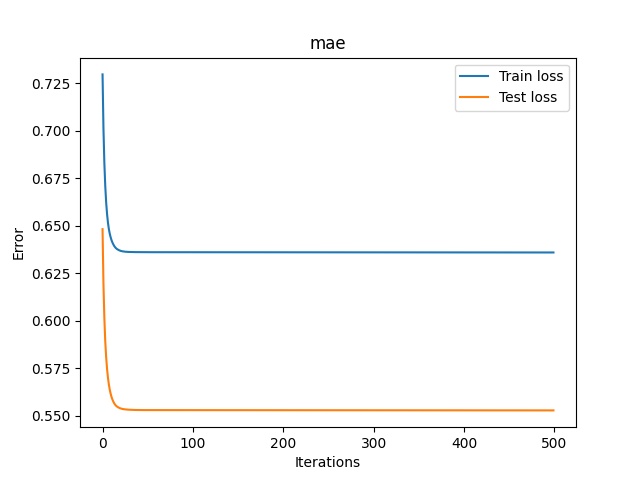
**Dataset-0**

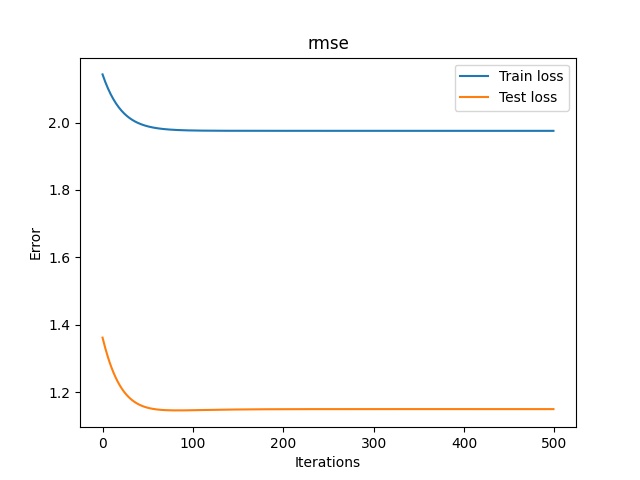
* Learning Rate=0.5, Iterations=500
* K=10
* Best MAE error on fold 9 = 1.5361122
* Avg MAE error with (K=10) = 1.7043154
* Best RMSE error on fold 9 = 2.18025189
* Avg RMSE error with (k=10) = 2.4393020765

****

****

**Dataset-1**

* Learning Rate=0.00001, Iterations=500
* K=10
* Best MAE error on fold 8 = 0.546223206495
* Avg MAE error with (K=10) = 0.6277619473
* Best RMSE error on fold 9 = 1.1499730746906007
* Avg RMSE error with (k=10) = 1.7836102244359928



**1d)**

1. [MAE]<=[RMSE]
2. RMSE and MAE are expected to have similar values in case of a low variance , when all the errors have same manitude.
3. MAE would be preffered in such a case as when MAE=RMSE the only thing we do is optimise the functions. And we know that MAE is harer to optimise as it is non-differentiable piecewise function. Also MAE is computationaly expensive.

**1e)**

*DataSet-1*

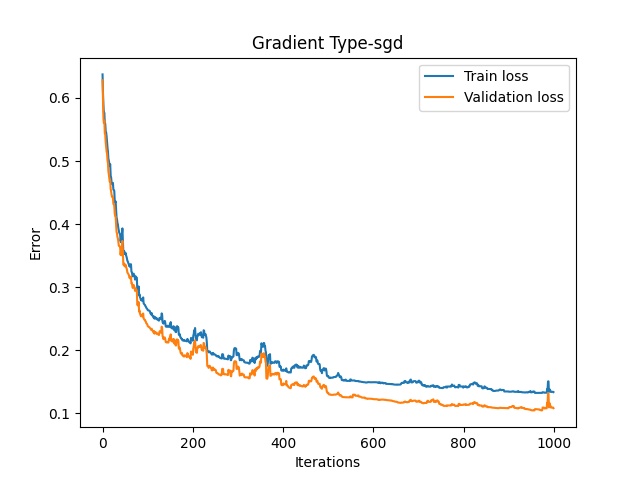
* Test Cost= 0.6826994305873468
* Train Cost= 0.77620795

*DataSet-0*

* Test Cost = 1.5482838865468944
* Train Cost = 1.641424797219717

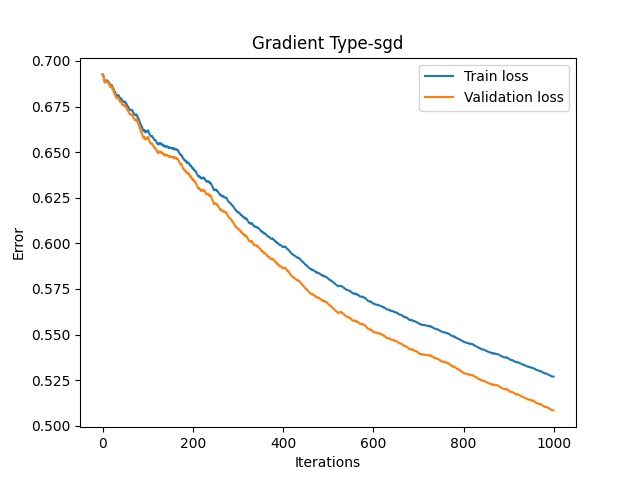
**Q2.**

**SGD**

* Learning Rate=0.01
* Iterations=1000
* Training Loss= 0.1339008657965337
* Test Loss= 0.14438876843579787
* Accuracy on Test Set= 93.43065693430657
* Accuracy on Train Set= 94.89583333333333

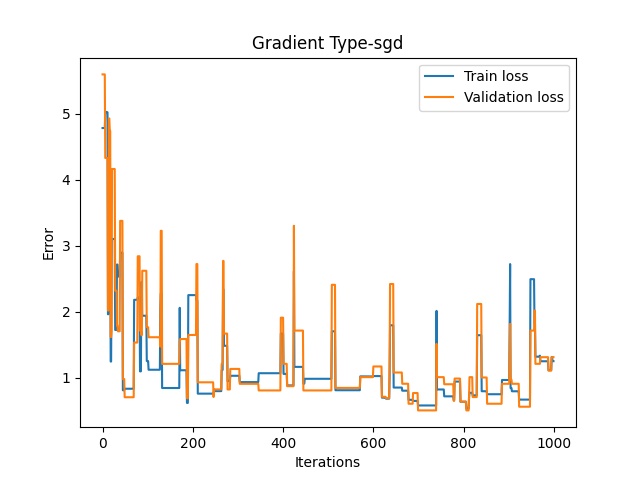
**SGD**

* Learning Rate=0.0001
* Iterations=1000
* Training Loss= 0.5247066673918643
* Test Loss= 0.5329540839051679
* Accuracy on Test Set= 73.72262773722628
* Accuracy on Train Set= 77.00312174817898



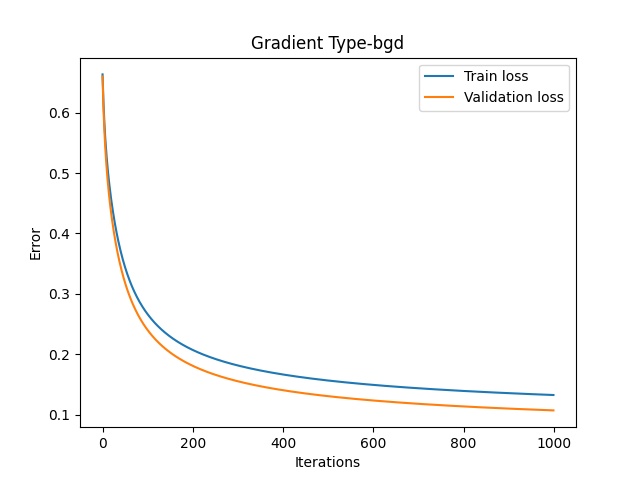
**SGD**

* Learning Rate=10
* Iterations=1000
* Training Loss = 1.250569919394072
* Test Loss= 1.5201584770735714
* Accuracy on Test Set= 88.6861313868613
* Accuracy on Train Set = 90.6347554630593

****

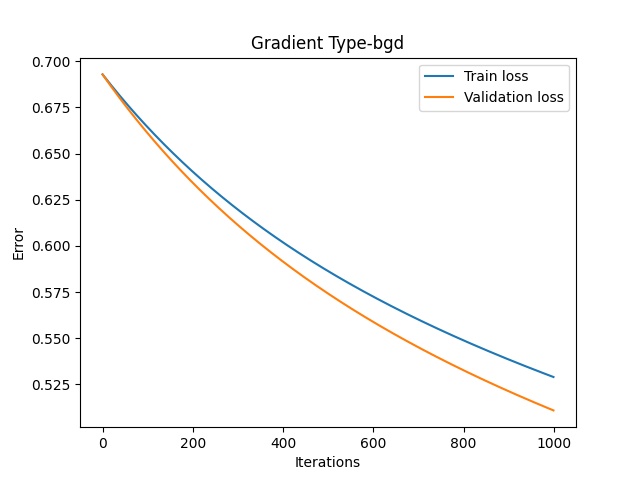
**BGD**

* Learning rate =0.01
* Iterations=1000
* Training Loss= 0.13258924925210067
* Test Loss= 0.1429357222216019
* Accuracy on Test Set= 94.16058394160584
* Accuracy on Train Set= 95.72916666666667



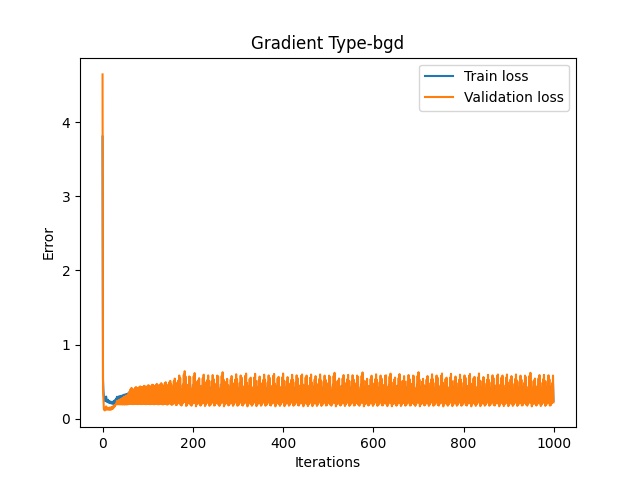
**BGD**

* Learning rate =0.0001
* Iterations=1000
* Training Loss= 0.5242444758577723
* Test Loss= 0.5330463619018994
* Accuracy on Test Set= 73.35766423357664
* Accuracy on Train Set= 75.96253902185224



**BGD**

* Learning rate =10
* Iterations=1000
* Training Loss= 0.2412454687975458
* Test Loss= 0.28605837782571913
* Accuracy on Test Set= 93.06569343065694
* Accuracy on Train Set= 93.75650364203955



* **Comparing Loss Plots🡪** The Sgd loss plot has spikes in the plot in between as for Sgd we have to randomly select a data sample and then update theta using that sample, we choose a random sample for every iteration.
* **Number of Epochs🡪** Sgd takes more number of epochs to converge as compared to Bgd.
* Sk-Learn Logistic Regression accuracy :-
  1. on Train data = 0.9937565036420395
  2. on Test data = 0.9817518248175182
* Sk-Learn SGD accuracy :-
  1. on Train data = 0.9864724245577523
  2. on Test data = 0.9817518248175182

**EDA**