Assignment 3 Report

Harsh Vora
CWID : A20445400
CS577 – S20

March 23, 2020

Problem Statement:

To tune the hyperparameters for two UCI datasets one for multi class classification and one for regression.
Evaluate different loss functions on 2 datasets.
Evaluate different optimizers on 2 datasets.
Evaluate the effect of different regularization measure on the performance of the classifiers on 2 datasets.
Summarize the result obtained in above experiments and draw conclusion based on the observation.

Implementation details:

I have taken UCI crime data repository for performing regression and UCI IRIS data for classification. I have run the hyperparameters tuning for various optimizers and losses first without normalization and then with normalization.

For Classification I have used Categorical Cross Entropy Loss, Kullback Leibler Loss and Hinge Categorical Loss.
For Regression I have used Logcosh loss, MSE(L2 Loss) & MAE(L1 Loss).
I ran the test 2 times once with normalization and other without normalization.
I have used Batch Normalization and Dropout during Normalization.
I have also done ensemble classification on IRIS Dataset.

Below are my list of hyperparameters for SGD, RMSProp, AdaGrad and Adam:

SGD = Learning Rate : 0.0001, Decay = 1e-6, Momentum = 0.9, Nestereov = True
RMS = Learning Rate: 0.0001, Rho = 0.9, epsilon = None, decay = 0.0
AdaGrad = learning Rate: 0.0001, Rho = 0.95, Epsilon = None, decay = 0.0
Adam = Learning Rate: 0.001, Beta 1 = 0.9, Beta 2 = 0.999, epsilon = None, decay = 0.0, amsgrad = False

For the regression model I have used the below network layers:
Dense Layer (64 neurons)
Activation = relu
Dense Layer (32 neurons)
Activation = relu Dense Layer (1 neuron)

For classification I have used the below network layers:
Dense Layer (32 neurons) Activation = relu
Dense Layer (16 neurons) Activation = relu
Dense Layer (3 neurons) Activation = softmax

Test Cases:

Regression:

Below are my results for UCI Crime dataset: I have measured the validation losses without normalization and with normalization:

https://archive.ics.uci.edu/ml/machine-learning-databases/communities/communities.data

Results without normalization:

Losses(Validation Loss)	Adam	SGD	AdaGrad	RMSProp
MSE(L2 Loss)	0.0217	0.0377	0.0199	0.0213
MAE(L1 Loss)	0.0919	0.1154	0.0974	0.1003
Logcosh	0.0089	0.0129	0.0098	0.0098

Validation Loss - Without Normalization

Results with normalization:

Losses(Validation Loss)	Adam	SGD	AdaGrad	RMSProp
MSE(L2 Loss)	0.0196	0.0551	0.0201	0.0252
MAE(L1 Loss)	0.0889	0.1394	0.0938	0.1010
Logcosh	0.0083	0.0197	0.0099	0.0099

Validation Loss - With Normalization

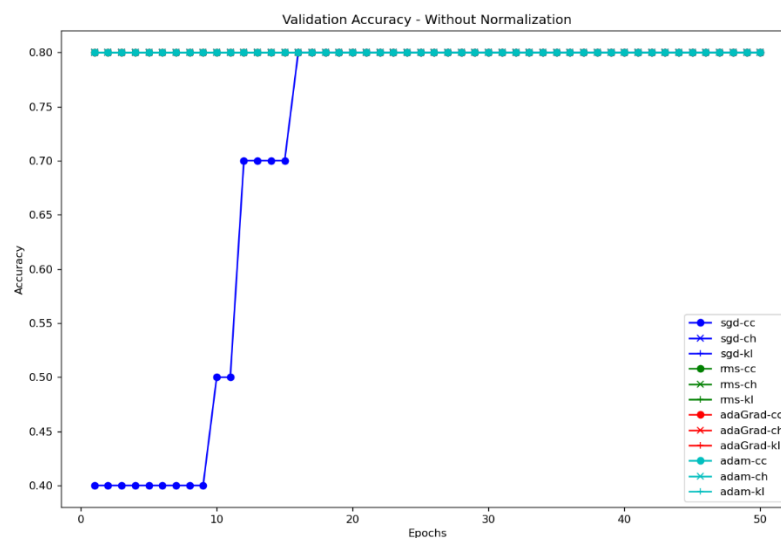**Classification:**

Below are my results for UCI IRIS dataset: I have measured the Accuracy without normalization and with normalization:

https://archive.ics.uci.edu/ml/machine-learning-databases/iris /iris.data

Results without normalization:

Losses(Accuracy)	Adam	SGD	AdaGrad	RMSProp
Categorical Cross Entropy	0.3579	0.6370	0.3819	0.4542
Hinge Categorical	0.3582	0.6700	0.5328	0.4812
Kullback-Liebler	0.3155	0.5204	0.3816	0.3819

Results with normalization:

Losses(Accuracy)	Adam	SGD	AdaGrad	RMSProp
Categorical Cross Entropy	0.9199	0.9288	0.6031	0.7324
Hinge Categorical	0.9399	0.8923	0.7073	0.7431
Kullback-Liebler	0.9200	0.8100	0.6100	0.5986

Validation Accuracy - With Normalization
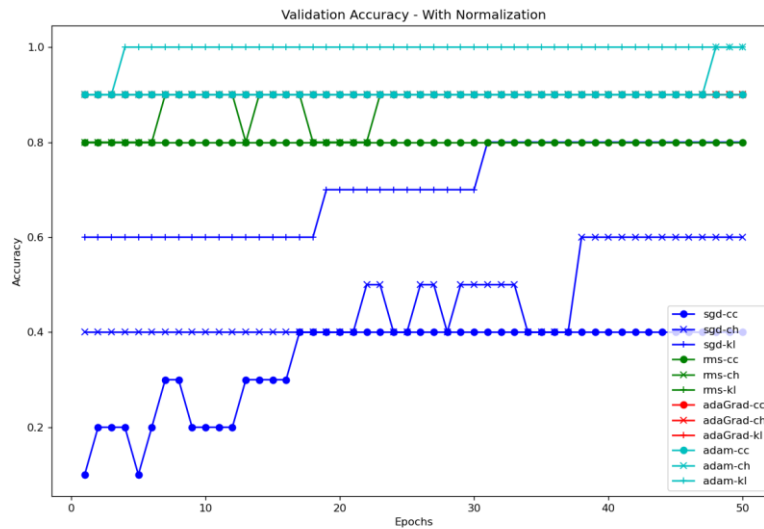
Ensemble Classification:

For ensemble classification I have used average of all prediction without normalization from all above models and taken an unweighted average.
Although I get maximum accuracy of 0.8 for Adam Optimizer. I am able to get a better test accuracy of 0.86 when using ensemble classification where I take the unweighted accuracy.

Conclusion:

My results for regression was best with Adam optimizer and logcosh loss when used Batch Normalization and Dropout.
For Classification I got the best results with Adam Optimizer and Categorical Cross Entropy Loss.
But I also noticed a higher accuracy obtained when I used batch normalization with Dropout in case of Adam, RMSprop and AdaGrad.
Without Normlization I obtained an accuracy of 0.9 using Adam and after normalization it is improved to 1.
So the best parameter set for classification problem was Adam using Categorical Cross Entropy and BatchNormalization with Dropout.