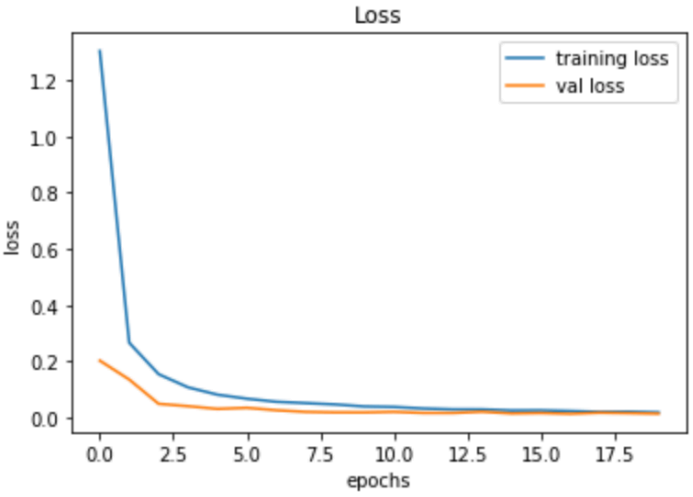
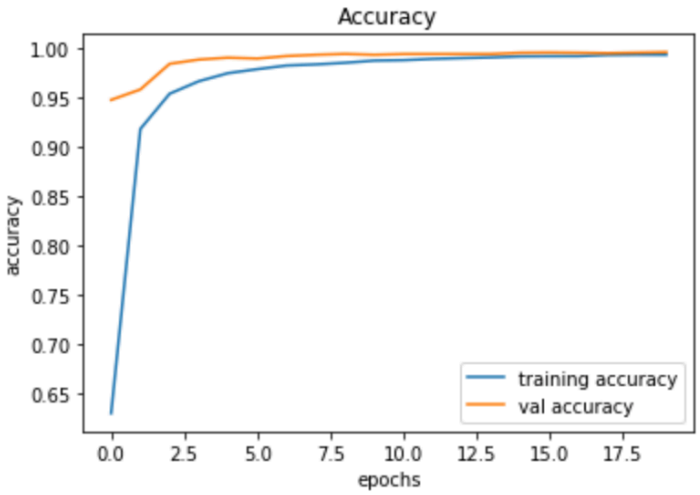
Assignment 8 Report
Tara Rasti

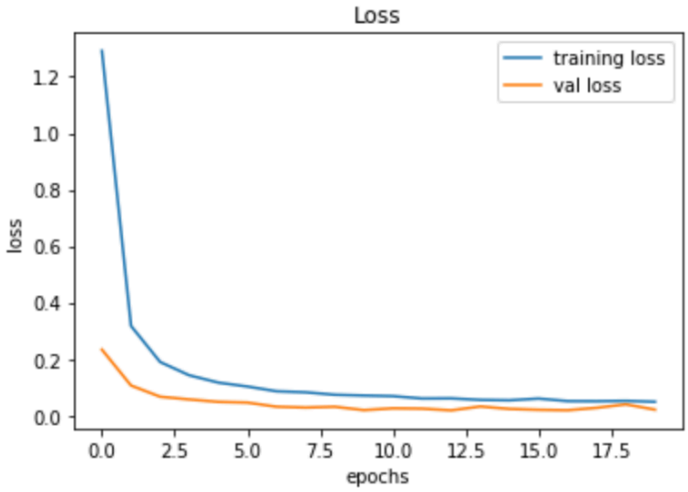
Classification

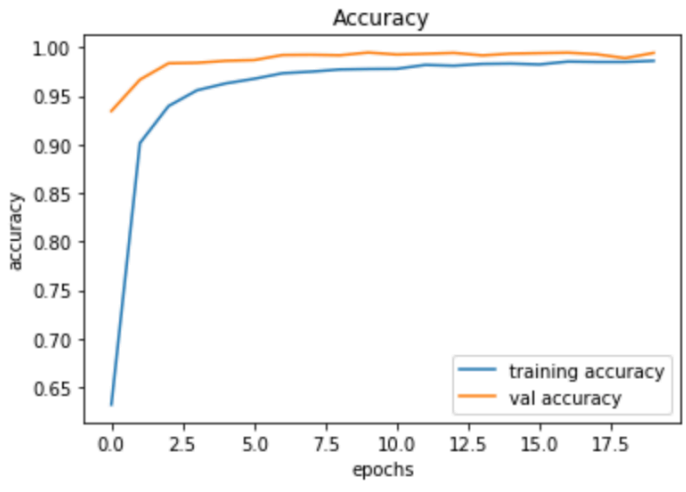
We have trained models by four different optimizers: Adam, Adadelta, SGD and RMSprop. Here we can see some of the interesting observations.
First you can see the loss and accuracy plots for a classification task. These plots are related to the RMSprop optimizer.

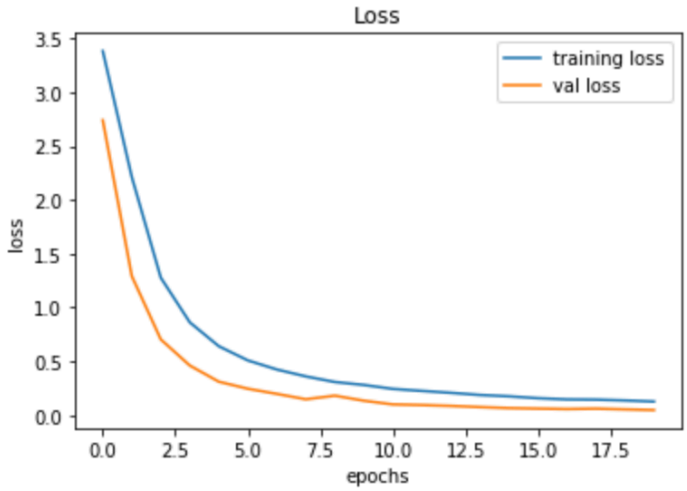


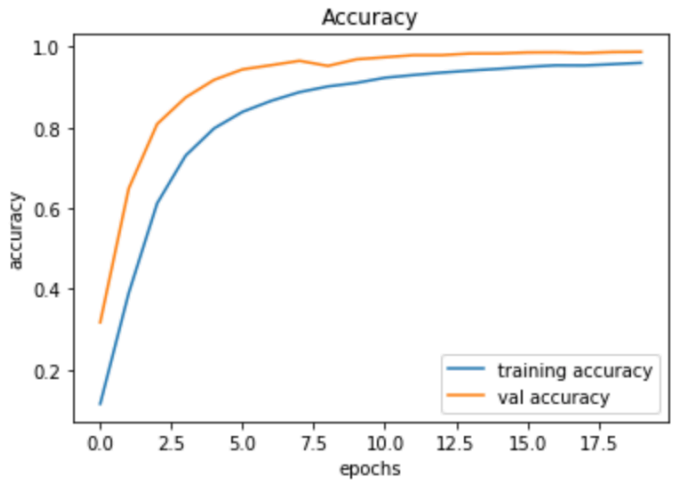


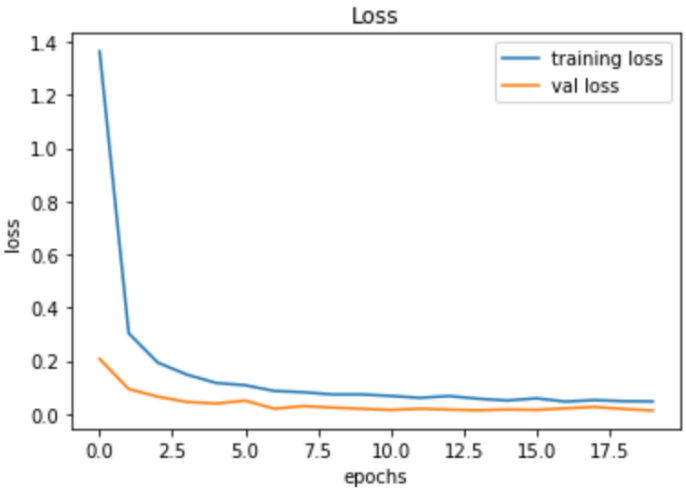
As you can observe, the RMSprop optimizer has acted very well and approximately after 15 epochs the train and val accuracy have become equal.
Below you can see the accuracy and loss plots related to the SGD, Adadelta and Adam optimizers respectively for this classification task.

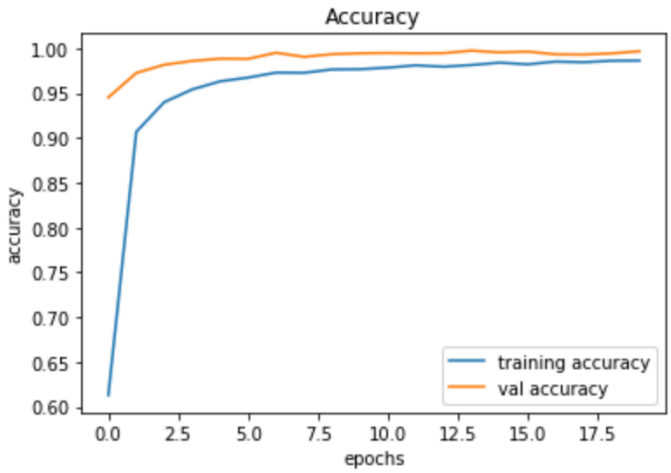










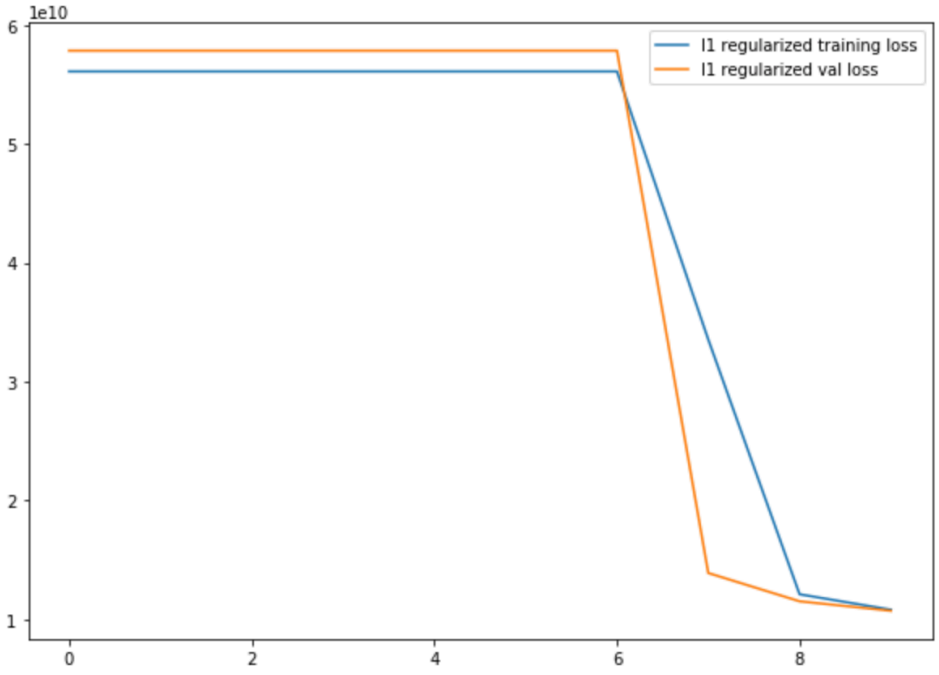


It is obvious from the plots that in RMSprop, Adam and SDG methods first there was a great gap between train and validation accuracy and also in train and validation loss. However, as number of epochs increases this gap decreases and in RMSprop method equals to 0. Nevertheless, Adadelta method acts nearly better from the first epoch.
Totally, in this case we can conclude that changing the optimizer does not change the results at the end as all of the methods ends to 99% accuracy for both training and validation sets.

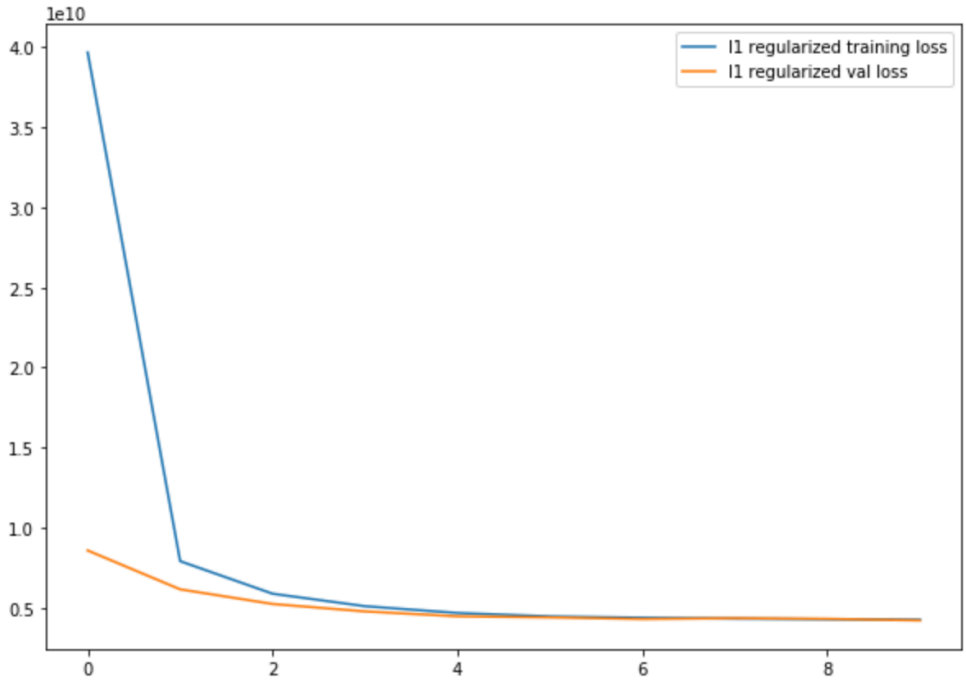
Regression

However there are some other cases in which there is a considerable difference between different optimizers. And the results change dramatically from NaN or constant to very good fit with small loss.
In the page below you can see the result of different optimizers(Adadelta, Adam, RMSprop and SGD respectively) on training loss in a regression task.

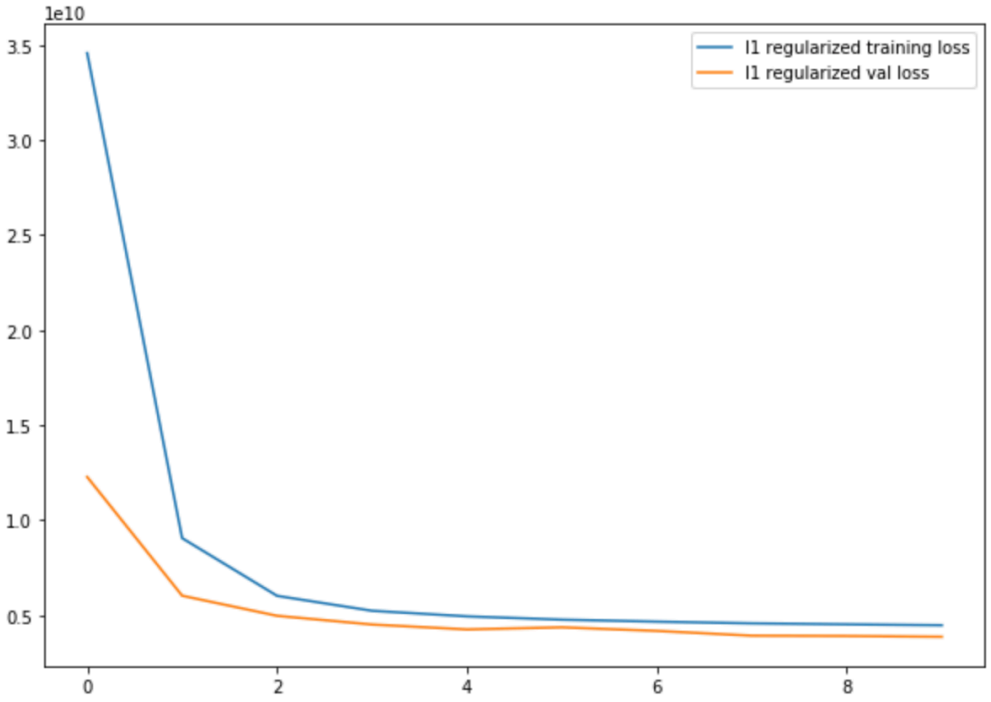
Adadelta



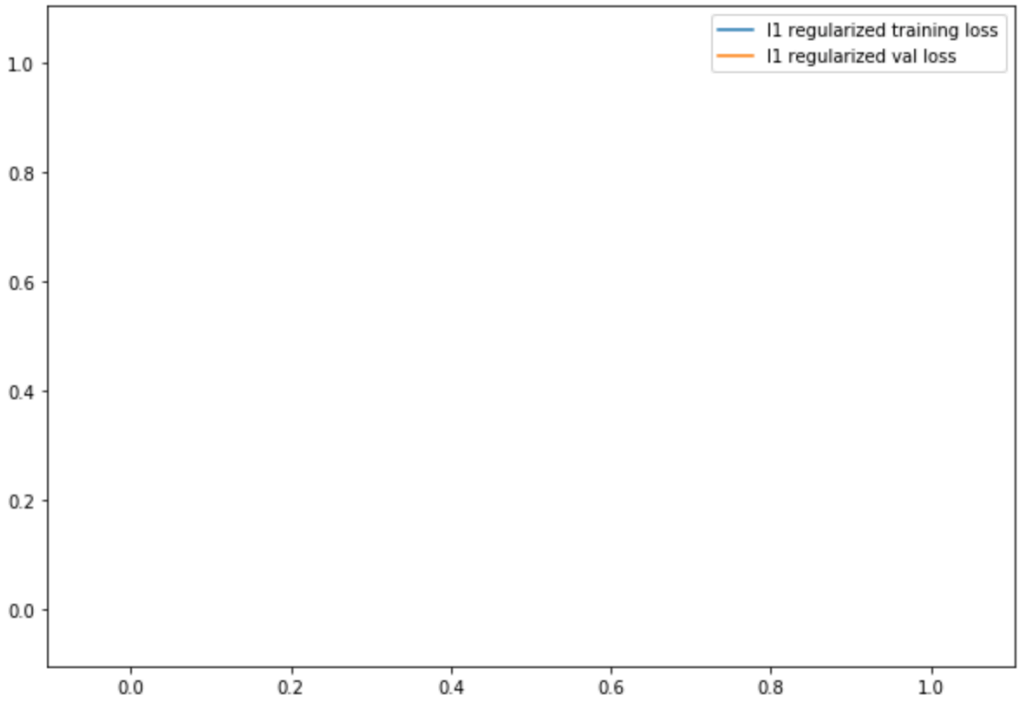
Adam



RMSprop


1e10
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
2
4
6
8
l1 regularized training loss
l1 regularized val loss

SGD


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
l1 regularized training loss
l1 regularized val loss