Logistic Regresion

Natalia Andrea Durán Castro
Universidad de los Andes
na.duran@uniandes.edu.co

Ana M. Velosa Orduz
Universidad de los Andes
am.velosa@uniandes.edu.co

Abstract

In this work the recognition of emotions is performed by machine learnig through the database FER2013, taken from kaggle, taking into account only the acknowledgment of smile and then the recognition of 6 other basic emotions, Angry, Disgust, Fear, Happy , Sad, Surprise and Neutral. For the above, loss functions and an optimizer were used to enter a model with the highest possible accuracy.

1. Introduction

Machine learning is about teaching computer systems to learn a certain task. This idea of teaching a machine has been complicated to understated because learning is an inherent human capability that is not easy to explain, let alone teach it to a machine.[1]
Explaining a bit further what does machine learning is, we can say that this algorithm doesn´t has to be programed step by step and also observes what is the performance obtained by the algorithm implemented and taking in to account the hyperparameters and the response obtained to improve the previous performance.[1]
To understand how well the machine learning approach it is working we must select a function of loss. This function quantifies between the result obtained and the annotation. Depending on the problem that you are trying to solve, there are different loss functions that are adjusted in the best possible way to the problem. On the other hand, we use the optimizer which shape and mold your model into its most accurate possible form by futzing with the weights. The loss function is the guide to the terrain, telling the optimizer when it's moving in the right or wrong direction.[2]
One of the most common used optimizer is the Stochastic Gradient Descent (SGD),because it's easy to implement and understand. This optimizer consists of the minimization of the decrease of the gradient. By doing this operation, we mark the path that the machine learning algorithm must take in order to minimize the function of loss. On the other hand, the speed with which the function of the machine is minimized is known as the learning rate. This value must be

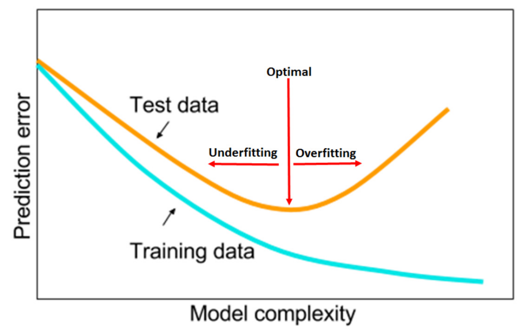

Figure 1. Model complexity vs. Predicion error expected curve

calculated in such a way that it can't be very small or very large, since the convergence of the algorithm that is being implemented depends on this value. In order to know if it is behaving as the optimizer should, the loss function is plotted as the optimizer is run, placing the number of iterations in the x-axis and in the y-axis, we put the values of the loss function. If the graph has a form like the one observed in 1, it means that the optimizer is working correctly. When the function stops decreasing it is because the cost function has converted to a minimum value.[2]

2. Methodology

For our emotion recognition detector we used the FER2013 data set from Kaggle. This dataset has a total of 31668,48x48 gray scale images divided in to 28079 train images and 3589 test images dividend in to 7 groups of images. Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral that have labels for each one of the emotions, from 0 to 6 as we can see in the figure 2. The train set was choosed as the first half of the original train set. In this way, the validation set was the other half of the train set.

The train.csv file its divided in 3 columns, the first column corresponds to the emotion label, the second one to the pixels and the third one to what group does the image goes to, either test or training. We divided training test in half to train and the other half to test the algorithm we are devel-

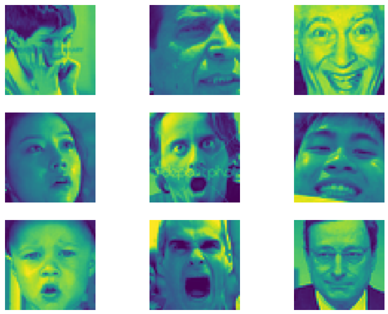

Figure 2. 9 Images form the DataSet

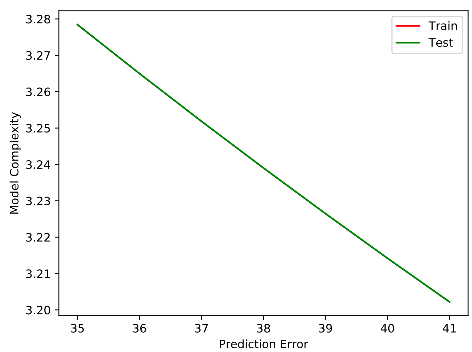

Figure 3. Model complexity vs. prediction error expected curve for the first 40 epochs in train and val

oping. The algorithm has to parts, the first part only detects smiley faces and the second one recognizes all 7 emotions. The two algorithms have the same structure, what change between one and the other is that the out when the algorithm is doing the model for the smile, because its one emotion its 1 and for all of the emotions its 7.

3. Results and discussion

The figure 3 shows the losses in the firsts 40 epochs. This value of epochs went when the loss in test set normalize to a value. Before this epoch, the values were infinites. On the other side, in the figure 4 is shown the epochs were the loss in train converge to a value. In the figures 5 and 6 are shown the results for a high number of epochs. The graphics have a similar behavior presents in the figure of the model 1. The behavior of both graphics can be classfied as optimal. Therefore, it is no underfitting or overfitting.

4. Conclusions

As expected it was obtained that in the graphics of Model complexity vs. prediction reaches a point at which its slope does not vary. This is because the values of the loss function

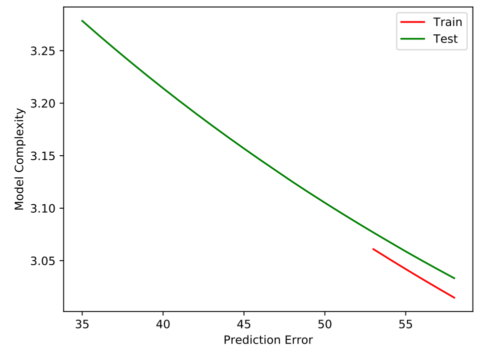

Figure 4. Model complexity vs. prediction error expected curve for the first 55 epochs in train and val

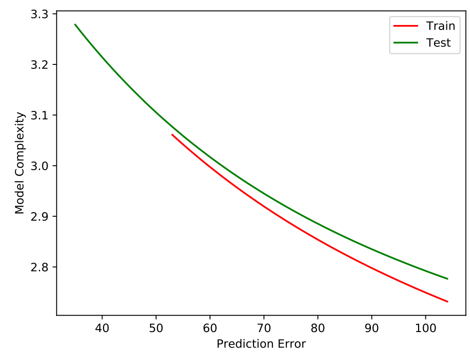

Figure 5. Model complexity vs. prediction error expected curve for the first 100 epochs in train and val

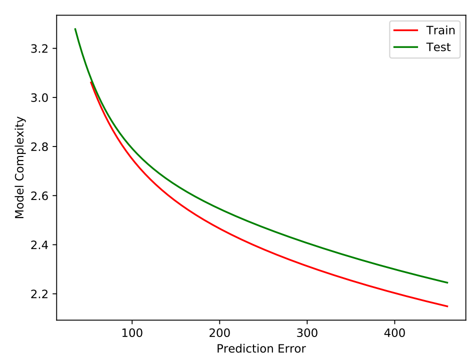

Figure 6. Model complexity vs. prediction error expected curve for the first 400 epochs in train and val

stop changing. In other words, the lost function converged to its minimum.

In addition to the above, it was not observed that the values of the loss function did not rise again, so it can be affirmed that a good learning rate was chosen and that a local minimum was not reached.

References

[1] R. Bali, D. Sarkar, B. Lantz, and C. Lesmeister. *R: Unleash Machine Learning Techniques*. Packt Publishing Ltd, 2016. 1

[2] S. Theodoridis. *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015. 1