# Facial Expression Recognition using Logistic Regresion

Luisa Fernanda Borbón R.
Universidad de los Andes
Departamento de Ingeniera Biomdica
lf.borbon@uniandes.edu.co

## Abstract

*Facial expression recognition is a challenging computer vision application that has multiple practical applications. Because of this, a basic Machine Learning approach was applied in the FER2013 human faces images in order to classify facial expressions presented on each image. The challenge was addressed using the Stochastic Gradient Descent as optimizer, first tackled as a binary problem with the Logistic Regression Loss Function to recognize happy faces, and then generalized to a multiple classification task using a Softmax Cross-Entropy approach.*

## 1. Introduction

From an early age, facial expressions are a basic way in which human beings communicate emotions, intentions and desires. Even thought recognizing this expressions is an easy task for humans, this can be pretty challenging for computers because images belonging to the same category often varies a lot from each other. This variation is not only explained by inherent differences between each picture taken, like changes in the lightning, brightness, pose or background; but also because each person show their emotions and facial expression differently [5].

Despite recent algorithms like artificial neural networks have gotten accuracies greater that 95%, this methods have been tested on easy frontal face and controlled environment images, and thus the task of recognizing emotions from more natural images remains unsolved [6]. Because of this, facial expression recognition presents as an interesting and challenging area of computer vision with multiple applications like mental state identification, security, lie detection, neuromarketing and other human-computer interaction systems.

In this sense, the present work intends to apply basic Machine Learning concepts in order to classify facial expressions presented on the FER2013 dataset. This was done using the Stochastic Gradient Descent (SGD) as optimizer and varying between the Logistic Regression loss function in the binary approach (Recognize happy faces) and the SoftMax loss function when having 7 multiple classification categories.

## 2. Materials and Methods

In order to recognize the emotions from FER2013 dataset images, as mentioned before the problem was first addresed . On the following sections, the dataset and the methodology followed are described in detail.

### 2.1. Dataset

The Kaggle Facial Expression Recognition Dataset was first released in 2013 on the International Conference on Machine Learning by Pierre-Luc Carrier and Aaron Courville. This dataset is composed by a total of 35888 grayscale 48x48 pixel images of human faces. For this project, the whole dataset was used, dividing its instances in 20000 training examples, 8709 validation images and 3589 test instances. All the images are annotated with on of the seven possible labels: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral and a sample of the images of the dataset is shown in figure 1[3] .

Figure 1. Example of human faces expressions form the FER2013 database [4].

## 2.2. Logistic Regression

The Logistic Regression analysis was used to tackle the binary approach of recognizing happy faces. This type of regression is used to describe and explain the relation between variables, according to a set of parameters. In this case, the sigmoid loss function is used to define the label of the image that is being classified, according to the way it fits in the regression parameters.

## 2.3. Soft-Max Regression

The multi-class classification was based on Dahal's Softmax and Cross Entropy Loss algorithm. In this approach, the sigmoid activation function is changed for a softmax function, shown in 1. This function takes an N-dimensional vector of real numbers and gives as a result other a vector with N real numbers ranging between (0,1) which add 1. Following this approach, for each image a 1x7 vector is created, getting on each position the probability of the image of being part of each one of the categories. Finally, the image is categorized as the prediction that had the higher probability [1].

$$p_i = \frac{e_i^a}{\sum_{k=1}^{N} e_k^a} \tag{1}$$

## 2.4. Stochastic Gradient Descent

As mentioned before, the Stochastic gradient descent (SGD) was used as the optimizing algorithm. This method adjusts the values of its parameters (w and b) in order to minimize a given cost function. Specially, the SGD algorithm does this minimization process updating the parameters for each training example within the dataset, one by one. By following the value of the gradient that minimizes the cost function, the algorithm can reach the values of w and b that achive the local minimun of the function [2].

## 2.5. Evaluation

To evaluate the accuracy of the algorithm, different metrics were applied over the test set. For each method used, the ACA, Precision and Recall Curve and the maximum F-measure were obtained.

# 3. Results

For the binary approach, the parameters were initialized randomly, varying the batch size and number of epochs taken into account. At the end, the best results were obtained using 200 training instances, with a learning rate of 0.0001 and which parameters were optimized over 40000 epochs. The loss vs epoch graph on figure 2, shows no significant evidence of over-fitting in the test set.

The precision and recall curve was obtained as shown in figure 3, verifying the predictions obtained with the algorithm in contrast to the actual annotations. In this algorithm, the F1 measure was 0.39 and the accuracy value was 0.48, which is low but expected as the approach taken is not complex enough to describe the data.

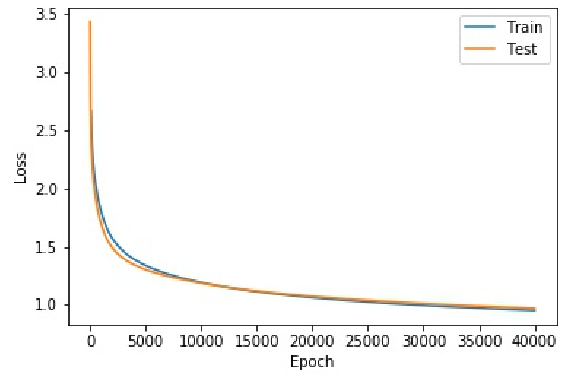


Figure 2. Loss value on the train and test subsets, according to the number of epochs.

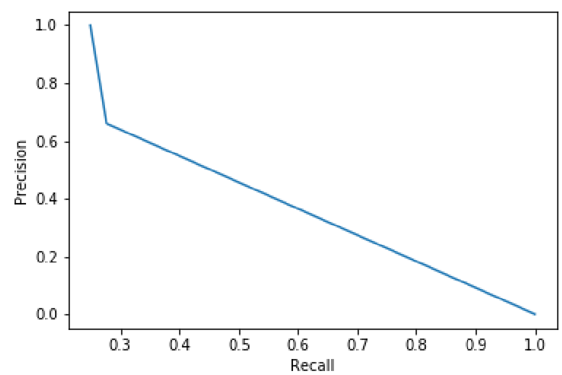


Figure 3. Precision and Recall curve achieved on the binary approach.

# 4. Conclusions

The chosen Logarithmic regression algorithm wasn't good enough to solve the facial expression recognition challenge. Looking at the images, it is really hard to distinguish between categories according to the feature representation space. Because of this, a richer representation could be taken into account including other descriptors for each image example.

# References


[1] E. Bendersky. The softmax function and its derivative, 2016. [Online] https://eli.thegreenplace.net/2016/the-

softmax-function-and-its-derivative/.

[2] N. Donges. Gradient descent in a nutshell, 2018. [Online] https://towardsdatascience.com/gradient-descent-in-a-nutshell-eaf8c18212f0.

[3] I. Goodfellow, D. Erhan, P.-L. Carrier, and Courville. Challenges in representation learning: A report on three machine learning contests, 2013.

[4] F. Knl. [deep learning lab] episode-3: fer2013, 2018. [Online] https://medium.com/@birdortyedi$_2 3820/deep-learning-lab-episode-3-fer2013-c38f2e052280$.

[5] A. T. Lopes, E. de Aguiar, A. F. D. Souza, and T. Oliveira-Santos. Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. *Pattern Recognition*, 61:610 – 628, 2017.

[6] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and vision Computing*, 27(6):803–816, 2009.