Logistic Regression as method for image classification

Saul Gómez
Universidad de Los Andes
sc.gomez11@uniandes.edu.co

Diego Valderrama
Universidad de Los Andes
df.valderrama@uniandes.edu.co

Abstract

Face expressions classification is a complex task in computer vision because of the high variations of the features from images of the same class and noise or artifacts from the environment. Logistic regression is classifier that learns to predict from the feedback obtained from its bad decisions and does not need a normally distribution of the data. The current study evaluates the performances of logistic regression in binary and multiclass face expressions classification problem in Fer2013 datasets. It was obtained an F-measure of 0.413 ACA of 0.55 for binary classification problem, Finally, the performance of our method in multiclass classification problem provides an ACA equals to 0.21.

1. Introduction

The improvement in image classification techniques have increased the possibility of assigning each pixel with an accurate class label [4]. However, such efforts still face some challenges, partly due to the high-dimension low-sample-size classification problem caused by the large number of narrow spectral bands with a small number of available labeled training samples [7]. Additionally, this problem, coupled with other difficulties such as high variations of the features from images of the same class, high similarities of spectral signatures between some different materials, and noise or artifacts from the environment, will significantly decrease the classification accuracy [7].

One approach to addressed this problem and obtain better results consists in apply logistic regression as a simple classification algorithm for learning to make the decision of classify a given image into a determined category, mainly in binary classification tasks. This results valuable to predict a discrete variable such as predicting whether a grid of pixel intensities represents one class to another. In this case, logistic regression uses a different hypothesis class to try to predict the probability that a given sample belongs to the "1" class versus the probability that it belongs to the "0" class [9].

The main advantage of using a logistic regression as a classifier is that it is not necessary for the data to be normally distributed, therefore it may be preferred when the data distribution is not normal, or the group sizes are unequal [1]. The main idea of structured sparse logistic regression model is to minimize the error function for classification with a sparsity constraint [**?**]. However, it is extremely sensitive to outlying responses and extreme points in the design space. Generally, logistic regression models are fit to data obtained under experimental conditions [6].

This approach has been studied in different applications for the task of image classification, for example, recognition of facial expressions. Now, being able to recognize facial expressions is key to non- verbal communication between humans, reason why the production, perception, and interpretation of facial expressions have been widely studied [8]. Due to the important role of facial expressions in human interaction, the ability to perform Facial Expression Recognition (FER) automatically enables a range of novel applications in computer vision [3]. The main idea relies on recognizing basic expressions, that is, expressions that convey universal emotions, usually anger, disgust, fear, happiness, sadness, and surprise, under naturalistic conditions. Nevertheless, this represents a challenging task due to variations in head pose and illumination, occlusions, and the fact that unposed expressions are often subtle [5].

In this same order of ideas, the current study presents an algorithm for facial expressions recognition by image classification from a logistic regression model fitted to the data. The problem is initially addressed as a binary classification task, followed by the analysis of the behavior of the method in a multi-class context, finally results are compared and discussed.

2. Methods

2.1. Database

FER2013 is a large, publicly available Facial Expression Recognition dataset consisting of 35,887 face crops with seven different basic expressions. The challenge of the dataset is is due to the fact that depicted faces vary significantly in terms of person age, face pose, and other factors, reflecting realistic conditions 1. The images in the database are distributed as follows: 4953 "Anger" images, 547 "Disgust" images, 5121 "Fear" images, 8989 "Happiness" images, 6077 "Sadness" images, 4002 "Surprise" images, and 6198 "Neutral" images. The dataset has not balanced categories, thus it is important to implement a robust method for these cases. Also, the dataset is split intro training, validation and test sets with 20000, 8709, and 3589 images, respectively. All images are 48x48 pixels grayscale. Additionally, the human accuracy calculated on this dataset is around 65.5% [2].

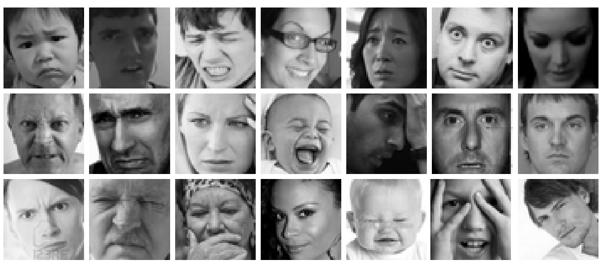

Figure 1. Example images from the FER2013 dataset, illustrating variabilities in illumination, age, pose, expression intensity, and occlusions that occur under realistic context. Images in the same column depict identical expressions with different intensity, such as anger, disgust, fear, happiness, sadness, surprise and neutral

2.2. Algorithm

The first approach was in binary classification in which we wanted to classify happy images (label 1) from the rest ones (label 0). Our method consists of a logistic linear regression as a classification method. For the algorithm training, we selected a small group of images (batch) and some labels were predicted. Then, with the predicted labels and the labels obtained from the data set for the analyzed images, we obtained the error function and the gradient (we initialized in a random way this values) in order to made a feedback of our model. Then, we iterated this procedure until all images of the training set were analyzed (1 epoch). After 1 epoch, the labels of the validation set were predicted and the error function associated with this data was obtained. Finally, this procedure was performed for a certain number of epochs.

After having established the general structure of our method, we performed some experiments to obtain the optimal hyperparameters (learning rate, epochs, batch). For do this, We ranged the size of the batch in 50,100,250 and 500 images with a learning rate of 0.001 and 40000 epochs. Then, we selected the batch size that minimized the error function. After that, we used 40000 epochs and the optimal value for batch size found in order to ranged the learning rate in 0.001 , 0.0001 , 0.00001 , 0.000001 and 0.0000001. Again, we selected the value that minimized the error function. Later, with the values of learning rate and batch size we ranged the number of epochs in 20K, 40K, 60K, 80K and 100K and selected the best hyperparamer. Then, we used precision-recall curve, F1 measure and ACA in order to evaluated our algorithm in test set.

After that, we addressed multiclass problem that consist in classify seven image classes: angry (label 0), disgust (label 1), fear (label 2), happy (label 3), sad (label 4), surprise (label 5), and neutral (label 6). To do this, we used seven binary logistic regressions with the optimal hyperparameters found in the binary classification problem. In this situation, each model was trained in order to identify a specific image class, i.e angry vs the rest ones or sad vs the rest ones. Then, given a test image, we predicted a probability with each model and we choose the label whose probability is maximum. Finally, we used ACA for measured the performance of our method.

3. Results

Table 1 shows train and validation error in each selection step of the optimal hyperparameters. We also found that the process time was very similar when we ranged batch size and learning rate, but when we ranged the number of epochs the process time increase significantly between experiments. As you can see, select the batch size and the learning rate was easy because of there are big different between the results. On the other hand, select the number of epochs was more difficult because the results are very similar. In the first instance, we though about selecting 60000 epochs since it has the lowest error values. Nevertheless, the processing time is significantly higher compared with 40000 epochs and the difference between the error is not significant, so we take 40000 epochs as optimal value. Fig 2 shows the error function of our method using the optimal values.

Table 1. Behavior of the hyperparameters

Batch size	Learning rate	Epochs	Train error	Val error
50	0.00001	40000	0.8681	0.8919
100	0.00001	40000	0.9613	1.0001
250	0.00001	40000	1.1094	1.1461
500	0.00001	40000	1.2861	1.3002
50	0.001	40000	0.7712	0.7830
50	0.0001	40000	0.7621	0.7713
50	0.00001	40000	0.8105	0.8343
50	0.000001	40000	1.4853	1.5103
50	0.0000001	40000	2.1803	2.2083
50	0.0001	20000	0.7675	0.7765
50	0.0001	40000	0.7374	0.7558
50	0.0001	60000	0.7312	0.7430
50	0.0001	80000	0.7278	0.7502
50	0.0001	100000	0.72582	0.74529

After selecting the optimal hyperparameters (50 as batch size, a learning rate of 0.0001 and 40000 epochs) we evaluated the performance of our method. Fig 3 shows the confusion matrix for binary classification problem in which we obtained an ACA of 0.55 . The results presented in this figure mean that is difficult for our method to classify images with happiness expressions but the model identifies in a better way the images in which there are not expressions of happiness. On the other hand, the precision - recall curve (see Fig 4) shows that the precision of our method is less than 40 % and the F-measure is 0.413. Although normally these types of metrics are used for detection problems, being a binary classification problem these results allow us to verify that the method has difficulties in the classification of facial expressions.

On the other hand, for multiclass classification problem, we found that the errors associated with the logistic regressions implemented are very similar and range between 0.78 and 0.76 (see Fig 5). Furthermore, Fig 6 shows the confusion matrix for multiclass classification problem with an ACA of 0.21.

4. Discussion

First, as shown in the figure 3, the confusion matrix shows that the method is highly accurate in identifying facial expressions that do not correspond to happy face, however its performance decreases in the recognition of smiles, increasing the amount of false positives detected (See also figure 4). This is because in the binary context, being a task of classification one-vs-all, one have a greater amount of negatives compared to the positive ones, so the classifier learns better the Features of the negatives, This is reflected in the accuracy of the classification.

On the other hand, according to the results obtained for the multiclass task the method implemented presents a performance above random classification, obtaining an 21% ACA. As can be notice in figure 6, the classifier performs better in the recognition of facial expressions associated with happy faces. This is because the parameters used for the multiclass method, were the optimal hyperparameters found in the binary context in the classification of happy-vs-all, this is why more precision by classifying these facial expressions is obtained. In addition, the increase in performance obtained in comparison to that observed in the binary task is due to the fact that a discrimination between the remaining classes is now taken into account, which formed a large set of negatives in the binary case. Therefore, the accuracy is increased because there is a decrease with respect to the false positives found as a result of the distribution of these in the remaining classes.

From the obtained results were identified the easiest and most difficult classes for the classifier which are the happy categories and fear and disgust, respectively (See figure 6). Examples of the classes mentioned above can be seen in the figure 7. As you can see from the illustrations 7(a) and 7(b), the facial expressions corresponding to the happy category present people with a big smile on their face, which allows the removal of features associated with the shape of the mouth and in some cases the eyes are more open, allowing for easy recognition. In contrast, expressions of fear and disgust showed an equally low performance. As shown in the figure 7(c), the expressions of fear show no characteristic features, so there is a tendency to confuse this category with any of the remaining six (See figure 6). On the other hand, in the case of the category of disgust, exaggerated expressions lead to confusion in the classifier with respect to the other categories. An example of this can be seen in the figure 7(d), where the features formed by the mouth of the people produce an error in the classification, associating these images to the happy category 6. It is also worth noting an overfitting of the method since there is a clear tendency to classify a Query image in the happy category in a large percentage of cases, as shown in figure 6.

Now, from the results shown in the figure 8, a qualitative analysis of the performance of the algorithm developed for the binary context can be performed. As can be seen from the different examples, the classifier has a low accuracy to determine whether an expression corresponds to happiness or not. This can be explained by an overfitting to the training data, as it is only able to recognize this emotion when the patterns in the facial features are very similar to the images of the training set. For this reason, the classification is wrong when the person makes particular gestures with the mouth, such as taking out the tongue. On the other hand, a trend can be seen in the classification of images with neutral expressions with a prediction of happy expression, which can be explained due to the threshold selected to establish the probability of being classified as 1 or 0. This leads to a drastic decrease in algorithm performance.

Finally, figure 9 shows the performance of the algorithm implemented on facial expression recognition in random natural images extra to the dataset. As can be notice, most of the facial expressions predicted corresponds to the real expression in the image. However, there is a mismatching classifying neutral expressions which often are confused with happy expressions and vice versa. This may be due to similarity in the features of the mouth and eyes of people in the image. In addition, expressions such as sadness are incorrectly classified in this case because the person in the picture has his eyes closed. Although some of its gestures represent happiness, the algorithm tends to associate the shape of the eyes with an expression of sadness, demonstrating another bias of the method implemented.

5. Conclusions

In order to improve the results obtained for binary classification problem, it is recommended to create a more robust classifier which combines some features and not only the label to learn to take decisions and predict labels in a better way. On the other hand, for multiclass problem it is recommended to think in a classifier that can learn the differences between all categories and not only between two of the categories in order to reduce the error and improve the results. Although in this approach we initialized the method in a random way, the inicialization of the method is also an important hyperparameter to think about.

References

[1] Q. Cheng, P. K. Varshney, and R. K. Tiwari. Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 3:491–494, 2006.

[2] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64:59 – 63, 2015. Special Issue on "Deep Learning of Representations".

[3] B. Martínez and M. F. Valstar. Advances, challenges, and opportunities in automatic facial expression recognition. 2016.

[4] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42:1778–1790, 2004.

[5] C. Pramerdorfer and M. Kampel. Facial expression recognition using convolutional neural networks: State of the art. *CoRR*, abs/1612.02903, 2016.

[6] D. Pregibon. Logistic regression diagnostics. *Ann. Statist.*, 9(4):705–724, 07 1981.

[7] Y. Qian, M. Ye, and J. Zhou. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 51:2276–2291, 2013.

[8] E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:1113–1133, 2015.

[9] S. University. Deep learning tutorial: Logistic regression. Online. Availible at: http://ufldl.stanford.edu/tutorial/supervised/LogisticRegression/, undated.

Images

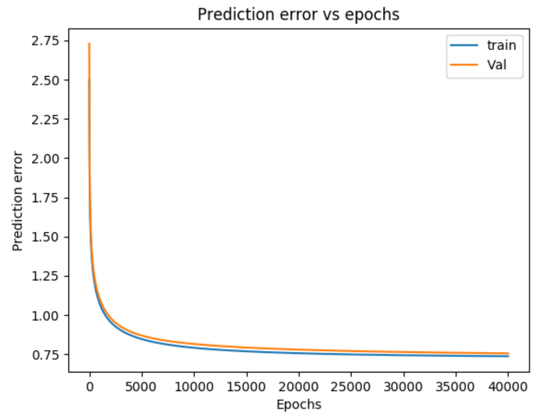

Figure 2. Logistic regression error in train and validation set for binary classification problem using the optimal hyperparameters. 0.7374 and 0.7558 are the minimum error in train and validation set after 40.000 epochs, respectively

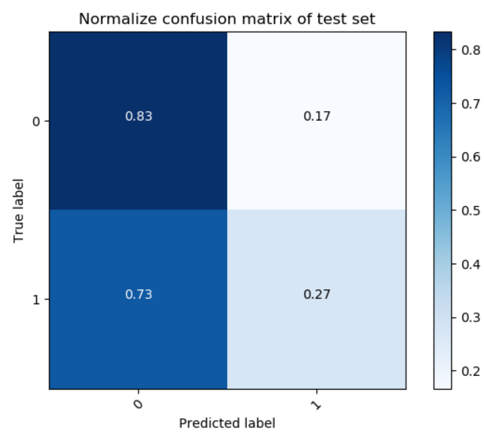

Figure 3. Confusion Matrix for binary classification problem. Label 1 means happy and 0 the rest ones. The ACA is 0.55

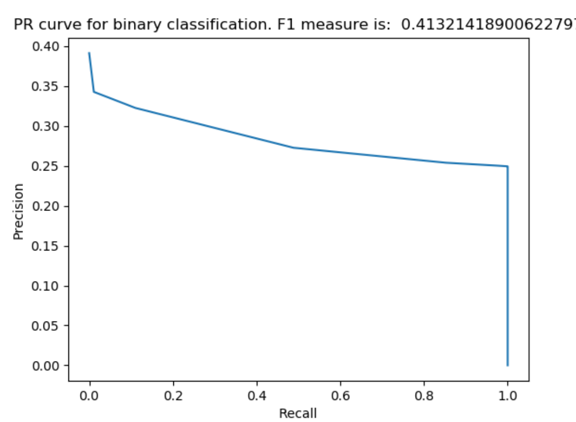

Figure 4. Precision-Recall curve for binary classification problem as detection problem. F-Measure is 0.413

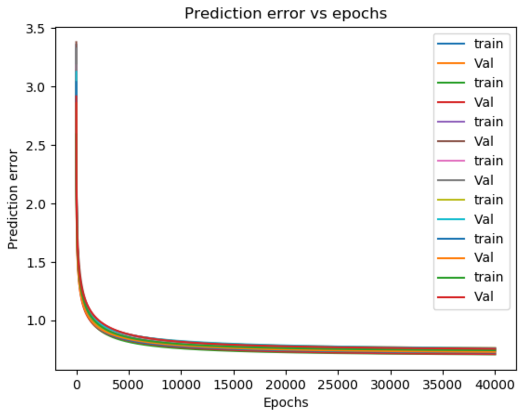

Figure 5. Logistic regression error in train and validation set for multiclass classification problem treated as 7 binary classification problems. The error values range between 0.77 and 0.75 for all problems.

(a) Happy expression

(b) Happy expression

(c) Fear expression

(d) Disgust expression

Figure 7. Examples of easier and harder facial expressions. Happy corresponds to the easier category for classification while fear and disgust represents the facial expression with lower performance in whole dataset

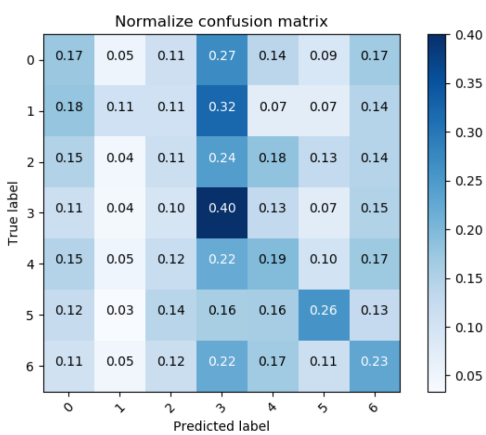

Figure 6. Confusion Matrix for multiclass classification problem. Label 0 is angry, 1 disgust, 2 fear, 3 happy, 4 sad, 5 surprise, 6 neutral. The ACA is 0.21

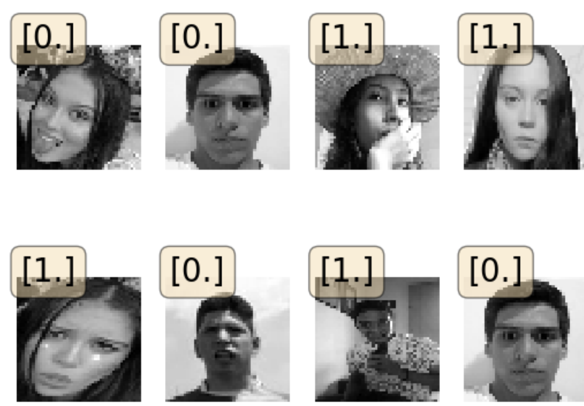

Figure 8. Performance of the method implemented in natural images external to the database. Label 0 is any facial expression except happiness and label 1 is happiness

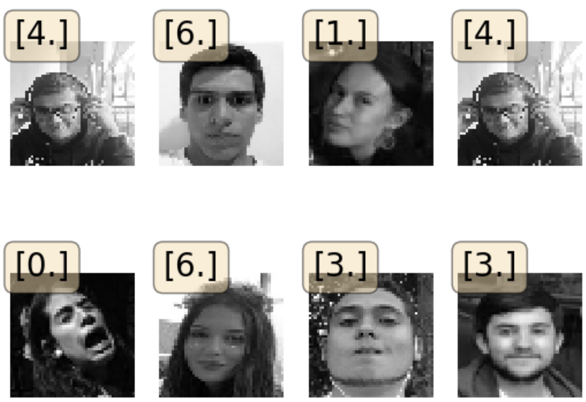

Figure 9. Performance of the method implemented in natural images external to the database. Label 0 is angry, 1 disgust, 2 fear, 3 happy, 4 sad, 5 surprise, 6 neutral.