Report Machine Learning Titanic Survived

by Bill Mono

1 Training a model

1. In my tests, I noticed that the best result in terms of precision of the model, i was able to obtain with a learning rate of 0.005.

2. Converges if we put a large number of iterations.

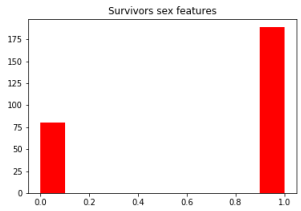

Fig. 1: Survive sex.

2 Analyze the model

1. My chances of survival weren't going to be very high. If I had been on the Titanic ship, my chances of surviving would have been around *27.1%* , due to the fact that in the test observation, I entered the following data. Ticket: 3, Gender: M, Age: 24, Siblings: 0, Parents: 0, I don't think I would have wanted to travel with them, and Fare: 7000. It is also due to the fact that to train my model I used the values of the ages and the learning rate which in training guaranteed me the highest accuracy, therefore 10000 and 0.0005.

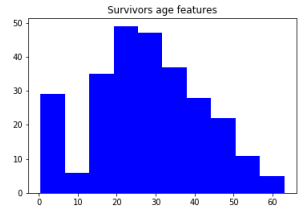

Fig. 2: Age feature survive.

2. The best accuracy of my model with training data was 80.4%, found by training the model with the fit function parameters set at 10000 and 0.005.

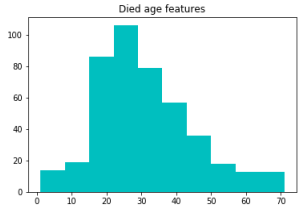

Fig. 3: Age feature died.

3. Looking at the parameters, what contributes to increase the chances of being saved is the 'sex' parameter, the others instead contribute to decrease the chances to survive.

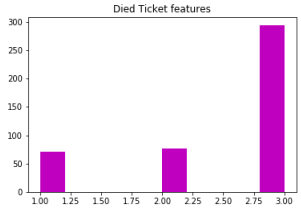

Fig. 4: Ticket feature.

4. As you can see from the graphs shown above, for the sex feature we can see that women are more likely to save themselves than men, as far as the feature ticket is concerned, here we see that people who are in the lower classes have a higher probability of not being saved than those with first or second class tickets. For the age I expected that the older people would have more difficulty saving themselves, and instead from the 2 graphs this difference is not marked, as regards the age it is seen that many children managed to save, compared to the number of those who are not saved.

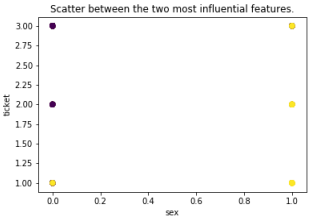

Fig. 5: Ticket feature.

5. The feature sex and ticket are in my opinion the most important ones that influence the model in one sense and the other. In this graph, females are saved whatever their ticket class is, while males are saved only if they are in first class.

3 Evaluated the model

1. The best accuracy of my model with training data was 79%, found by training the model with the fit function parameters set at 10000 and 0.005.

2. No, because the values of the accuracy of the training and the test are very close, obviously that of the training is higher but not such that we have an overfitting.

3. One way would be to change the validation method, making a random split on the data, the model is very affected by learning, of the data that is provided to it, and therefore depending on the data on which you will train its accuracy could be very different. Instead of using a random split, for example, we can use Cross Validation. Obviously another way to increase performance would be to have more data.

Note: I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher