Project A - Problem 1

Camelia D. Brumar

October 27, 2020

Contents

1 1A: Dataset Exploration **1**
1.1 Table 1A 1
1.2 Short Answer 1B 1
1.3 Answer for Figure 1c 1
1.4 Short Answer 1D 2
1.5 Short Answer 1E 3

1 1A: Dataset Exploration

1.1 Table 1A

	train	valid
total number of examples	9817	1983
number of positive examples	4913	1036
fraction of positive examples	0.5	0.522

Table 1: Summary of some basic properties of the provided training set and validation set: total count, positive label count and fraction of positive labels.

1.2 Short Answer 1B

Discuss the results you are seeing; what do they show, and why?

Answer: On one hand, the training BCE and training error rate monotonically decrease as the number of iterations increases. On the other hand, the validation BCE and validation error rate decrease as the iterations increase up to 10 iterations. However, for the models we fitted with more than 10 max iterations, the validation BCE increases dramatically as the number of iterations increases, and the validation error rate starts increasing moderatelly as the number of iterations increases.

This phenomena we're seeing in both plots, where after the max_iter=10, the training and validation errors split, one decreasing and the other one increasing, respectively, is called overfitting. Also, If we would choose max_iter to be less than 7 or 8, we would underfit our model.

1.3 Answer for Figure 1c

Which hyperparameter should you select?

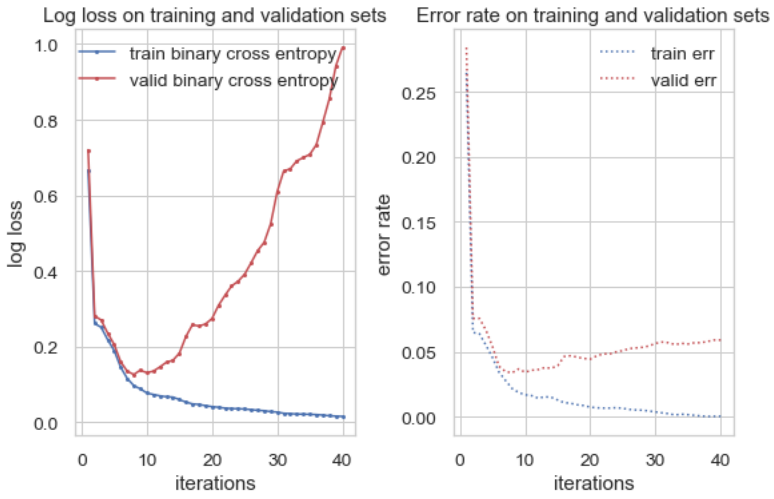

Figure 1B: BCE and Error rates on validation and training sets.

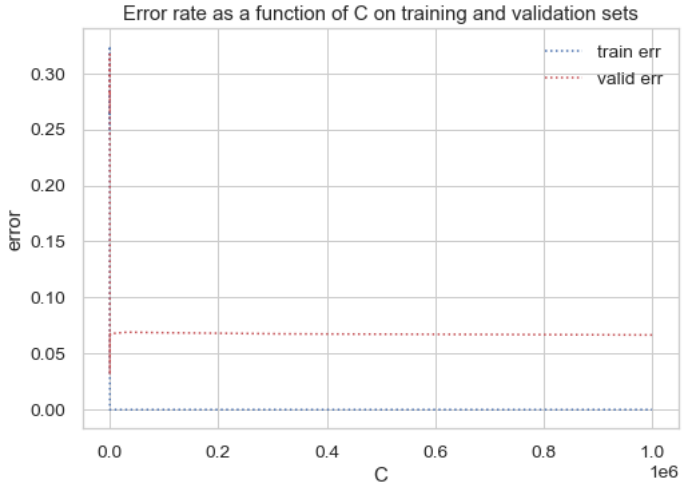

Figure 1C: Error rate as a function of C

Answer: The hyperparameter C I should select is the one that minimizes the validation error rate, in this case it's C = 0.01.

Note: I slightly modified the required figure 1c so the plot can be more readable. This does not affect which is the C that minimizes the validation error rate. See Figure 1C Improved.

1.4 Short Answer 1D

Discuss the results you are seeing in Figure 1D. What kinds of mistakes is the classifier making?

Answer: Seeing the results from Figure 1D, I realize that the classifier I chose is doing a bad job at predicting the label when the numbers are slightly rotated. Also, it doesn't predict well when an eight resembles a nine because the lower part of the number is very narrow. An example of this would be

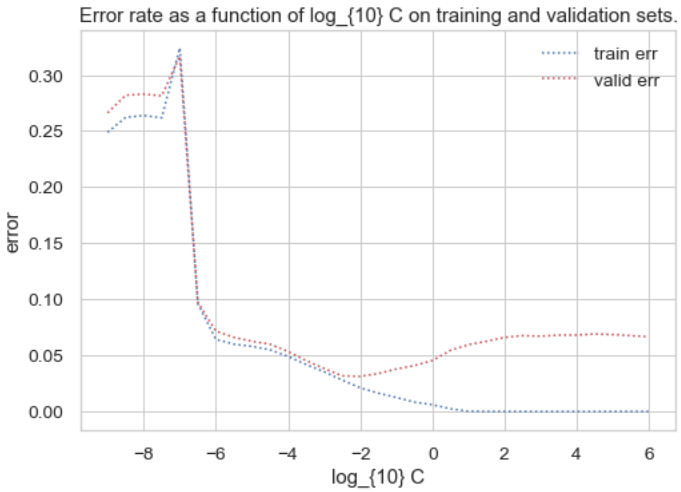

Figure 1C Improved: Error rate as a function of log_10(C)

Figure 1D - FP: 9 sample images that are false positives on the validation set.

the top left image in the false positives figure.

1.5 Short Answer 1E

Which pixels have negative weights, and thus have high-intensity values correspond to the negative class ('8')?

Answer: The pixels that are colored with colors between light yellow and intense red.

Which pixels have positive weights, and thus have high-intensity values correspond to

Figure 1D - FN: 9 sample images that are false negatives on the validation set.

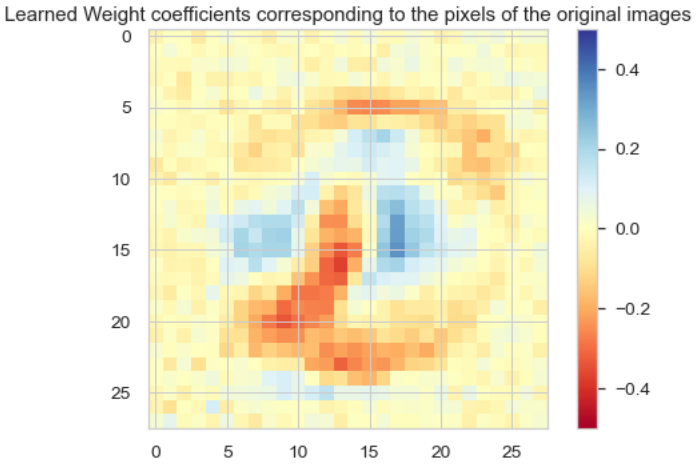

Figure 1E: Weights of the trained model represented as a heat map.

the positive class ('9')?

Answer: The pixels that are colored with colors between white and intense blue.

Why do you think this is the case?

Answer: Because the model learned the positions of where 8s lay on the image, i.e. the where the pixels are "on" when the image has label 0, and assigns them negative weights. The same happens with the pixels that represent the class 1 (the 9s), but now using positive weights.