

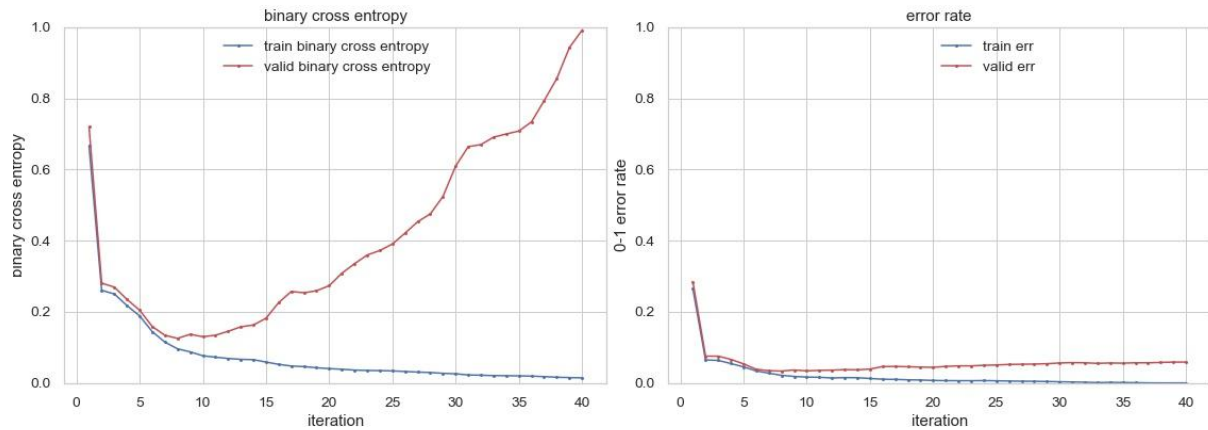
# Project A Report

## Problem 1

**Table 1A**

	Training Set	Validation Set
Total number	9817	1983
Number of positive	4913	1036
Fraction of positive	0.5004	0.5224

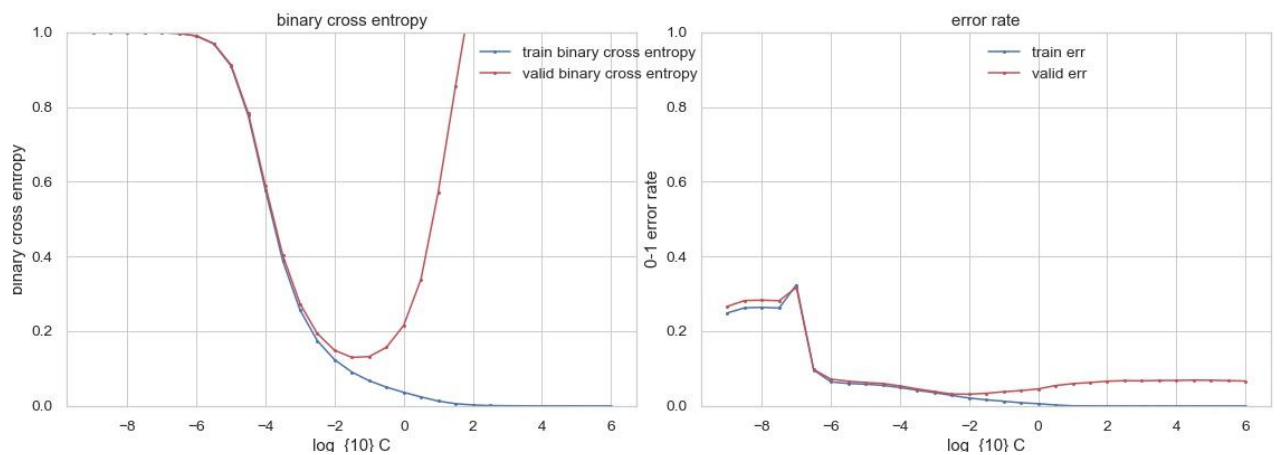
**Figure 1B**



Short Answer 1B : Below the plots, discuss the results you are seeing; what do they show, and why?

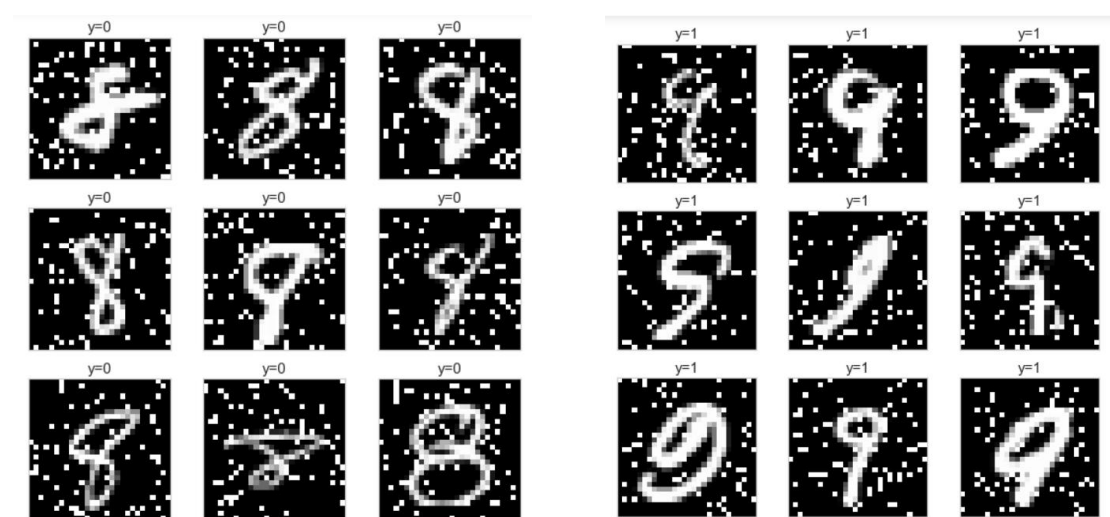
According to the plots, more iterations do not mean better results, because too many iterations can result in over-fitting to training data and has very bad performance on validation data, we should choose a proper number of iteration. We can see from the graph, 8 iteration have the best validation log loss, so we should choose some numbers around 8.

**Figure 1C**



According to the figure, I should choose  $C = 0.1$ , it give best log loss and error rate.

**Figure 1D: left contains 9 false positive, right contains 9 false negative**

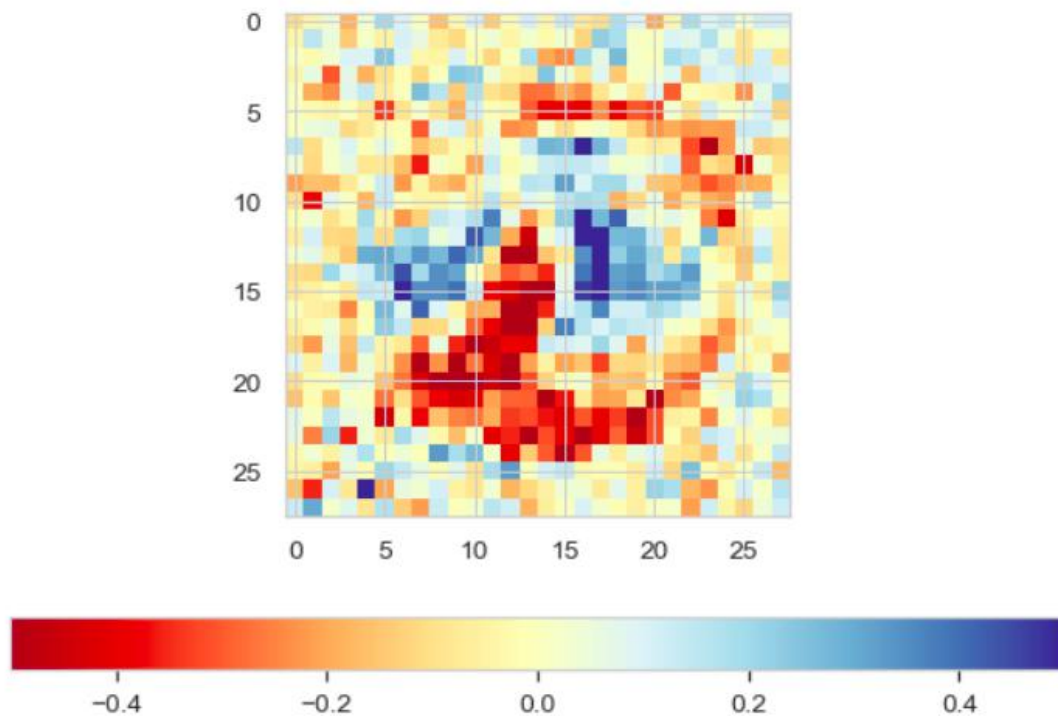


Short Answer 1D: Discuss the results you are seeing in Figure 1D. What kinds of mistakes is the classifier making?

Some of the mistakes are reasonable, like the 5<sup>th</sup> and 6<sup>th</sup> image in left plot, the 7<sup>th</sup> image in the right plot.

The mistakes the classifier makes is that it can not recognize number in strange shape, some of the handwritten number is not so standard and well written, but we can still tell it is 8 or 9 by its shape, but the classifier could not tell the number. And it seemed to be recognizing certain pixels rather than the curve and how it goes, so a strangely written number may be classified wrong.

**Figure 1E: weights of each pixel, and the mapping between color and weight**



**Short Answer 1E**

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

Which pixels have positive weights, and thus have high-intensity values correspond to the positive class ('9')?

Why do you think this is the case?

Pixels with yellow to red color have negative weights, it is obvious around bottom-left and top-right part with deep red color.

Pixels with blue color have positive weights, we can see these blue pixels are mainly around the center.

The reason is by looking at the colors, if a pixel have negative weights and it is white, it will contribute negative value to the total score, and pixel with positive value will contribute positive value to total score.