Since the possible values of are and . We apply the sigmoid function to the new dataset for .

We can write the form of probability as following (we suppose :

But we can write those more compactly as:

because .

Thus, we got:.

Similarly from how we got formular (1), apply the log loss here to get:

.

Then we apply the regularization to the loss function, and we want to minimize the negative log loss with the penalty. Thus, we got:

.

We got exactly the same objective as (2) in the question.

The difference between the two objectives (1) and (2) in this question is: for , , but for , . Except that, how we maximize the total probability, how we apply the negative log-loss and what the penalty is, are all the same. It is just the difference between the expression of probability, the essence after it is the same, we are counting the probability of two classes. It is just because the two classes’ values which we are classifying are different. That is why the two objectives are identical.