ITE4053 - practice2 report

Junyeong Park

1 Methods

In these experiments, we used dataset $\{x_i, y_i\}_{i=1}^{N}$ consisted of one-dimensional inputs and one binary outputs. For all i, x_i is a real number in $[0, 360]$. And y_i is 1 if $sin(x_i/2\pi) > 0$, else 0.

Like prevision practice but not exactly same, we use the logistic regression classifier to solve this problem. In this case, our model just required one input, not two inputs. We trained the models with gradient descent method. The hyperparameter α used in this practice are 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 2, 4, 8, 10, 15, 20.

2 Results

2.1 Estimated parameters

The best parameter that we got in these experiments is $w \approx -0.0059$, and $b \approx 0.2758$. Figure 1 represents the output of the logistic regression model.

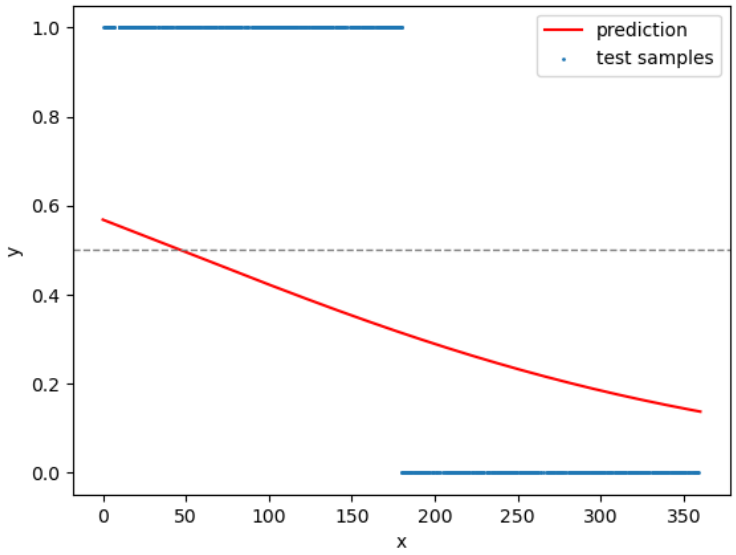

Figure 1: The best model for these experiments.

2.2 Best hyperparameter α

To get the best hyperparameter, we tested various candidates. In Fig. 2, the best hyperparameter α is 0.0001 that has the lowest error in test dataset.

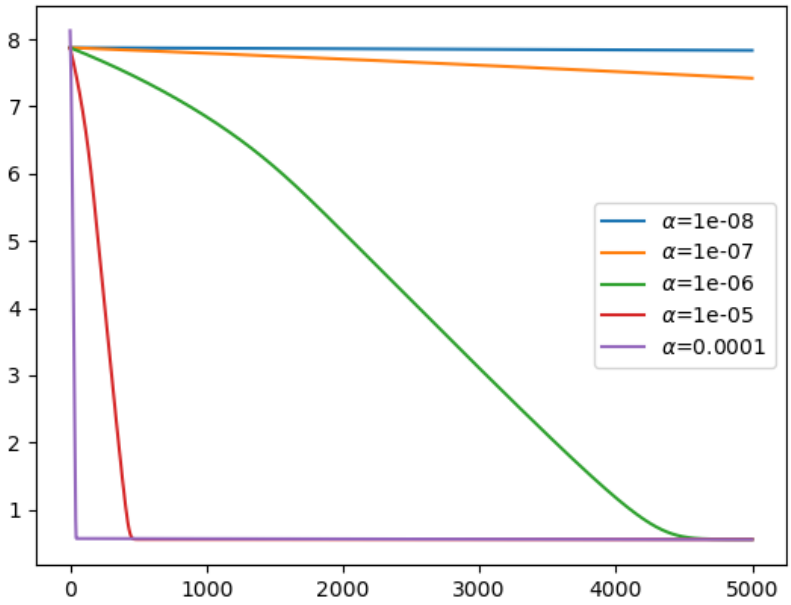

Figure 2: Model training performances for each hyperparameter. Some hyperparameter results are omitted because they are too noisy.

2.3 Accuracy

We frozen random seeds to reduce noise that is not an associated variable of the experiments. Below results are from the same datasets and the same initial parameters.

	(10, 1000, 5000)	(100, 1000, 5000)	(10000, 1000, 5000)
Accuracy (training set)	70.0	64.0	62.6
Accuracy (test set)	58.3	59.5	61.4

Table 1: Performances according to the number of updates. (m, n, K) denotes that the model was updated K times and that m training samples, and n test samples were used.

	(10000, 1000, 10)	(10000, 1000, 100)	(1000, 1000, 5000)
Accuracy (training set)	50.4	58.9	64.6
Accuracy (test set)	47.2	56.8	65.8

Table 2: Performances according to the number of training samples. (m, n, K) denotes that the model was updated K times and that m training samples, and n test samples were used.

3 Discussion

In Fig. 2, test samples are able to be linearly separable, so that it seems to be solvable with logistic regression. However, our logistic regression model has poor performance at this problem. This seems to be due to the large range of inputs. In previous practice, the range of the input is 0 to 1, but in this task, the range of input is 0 to 360. This is 360 times wider. In fact, when the range of inputs was reduced from 0 to 2π, the model was classified with high accuracy. Thus, the normalization of the input data was very important for the performance of the model.