report_lab3

Yanjie Lyu, Yi Yang, Qingxuan Cui

2024-12-11

Contents

1 Theory

1.1 In neural networks, what do we mean by mini-batch and epoch?

mini-batch: mini-batch is the subset of training data randomly sampled, aiming to solve the problem of excessive computation time and memory space consumption caused by using the entire training set to train the model and updating the parameters, especially when many of them are probably relatively similar data points in the training set. A mini-batch can typically contain 1, 10 or 100 data points, but it is recommended to set it to the nth power of 2, so that it can be aligned with the memory size of the GPU. epoch: One complete pass through the training data is called an epoch. An epoch consists of $\frac{n}{nb}$ where n represents the size of train data, and nb represents the batch size. (P124-P125)

2 Assignment 3 Support Vector Machine

filter	filter0	filter1	filter2	filter3
error_rate	0.165	0.167	0.150	0.014

2.1 Which filter do we return to the user ? Why?

Filter2 should be returned to the user.

Reasons:

Filter2 was trained by the union of validation data and train data, evaluated by test data, with the risk that the model overfitting the distribution of hyperparameter tuning. However, filter2 has the lowest error rate of any model evaluated using a dataset that did not appear in training at all.

As for the others: Filter0 was evaluated by validation data, while validation data have been used in optimization. The performance of the model may be overestimated. It may "memorize" the distribution of valid data, but perform poorly on real independent test data.

Filter1 was trained by train data and evaluated by test data. It is the standard pipeline, but the error rate is higher than filter2.

Filter3 was trained by the whole data set, which contains the test data set. After that, the model have adapted the test data in some degree. Thus, it would affects the accuracy of model evaluation.

Answer 2:

Filter3 should be returned to the user.

Reasons:

Filter3 is trained using the entire dataset (`spam`), which includes the training set, validation set, and test set. By leveraging all available data,Filter3 is able to utilize the maximum amount of information to build the most robust model possible.

Once the testing phase is complete and the model has been evaluated, the focus shifts to creating the best model for practical use. Using all available data to train this model is the optimal choice.

2.2 What is the estimate of the generalization error of the filter returned to the user? Why?

Error2 should be returned to the user. We used the filter2, because it perform better on new dataset among the models who were evaluated on new dataset. Discussion required there.

Answer 2:

This ensures that the evaluation is unbiased and reflects the model's ability to generalize to unseen data. Filter2's error is a more accurate estimate of the generalization error because the test set remains independent and has not been contaminated by the training process.

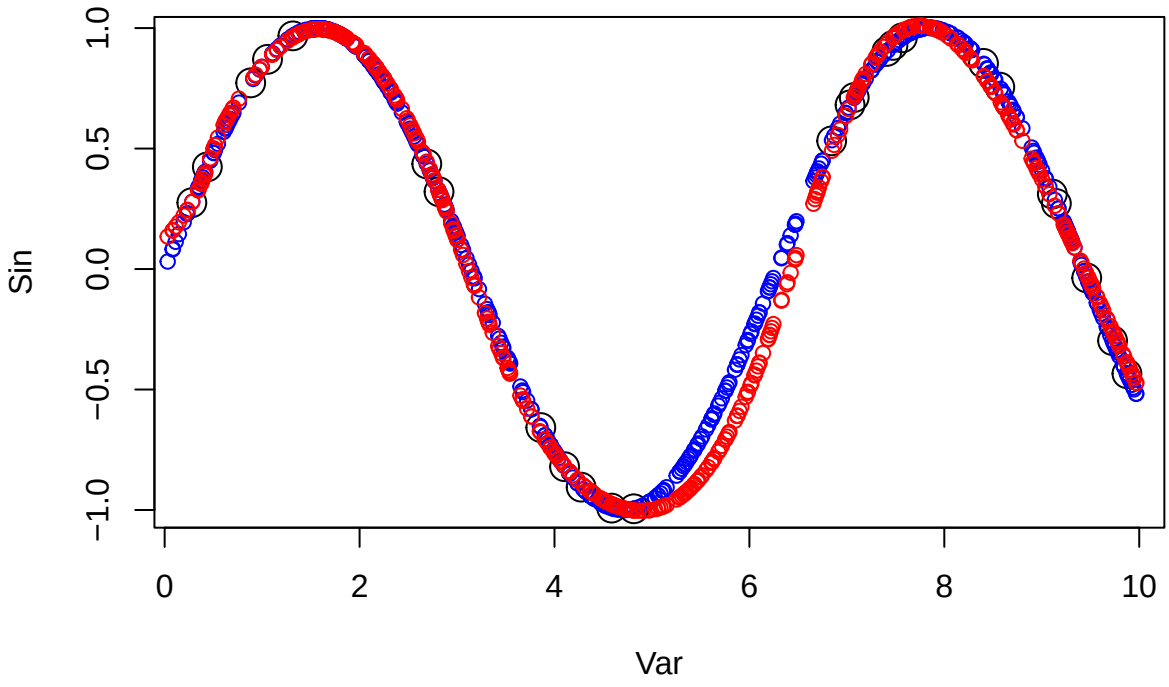
2.3 Implementation of SVM predictions.

Based on the table below, we can see the values and labels from manual prediction and function based prediction. Except the first data point, the labels are same.

```
table_pred = data.frame(pred_vec_label, pred_func_label)
colnames(table_pred) = c("manual prediction label", "prediction label")
row.names(table_pred) = paste0("data point", as.character(1:10))
kable(t(table_pred), format="markdown")
```

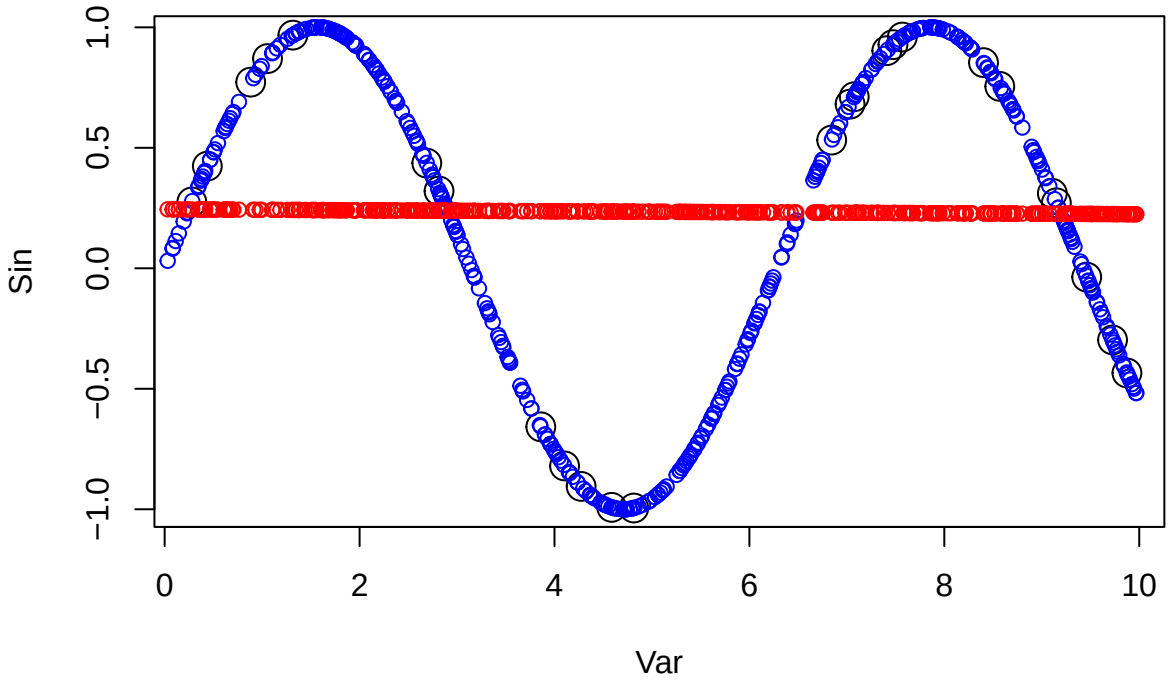
3 Assignment 4

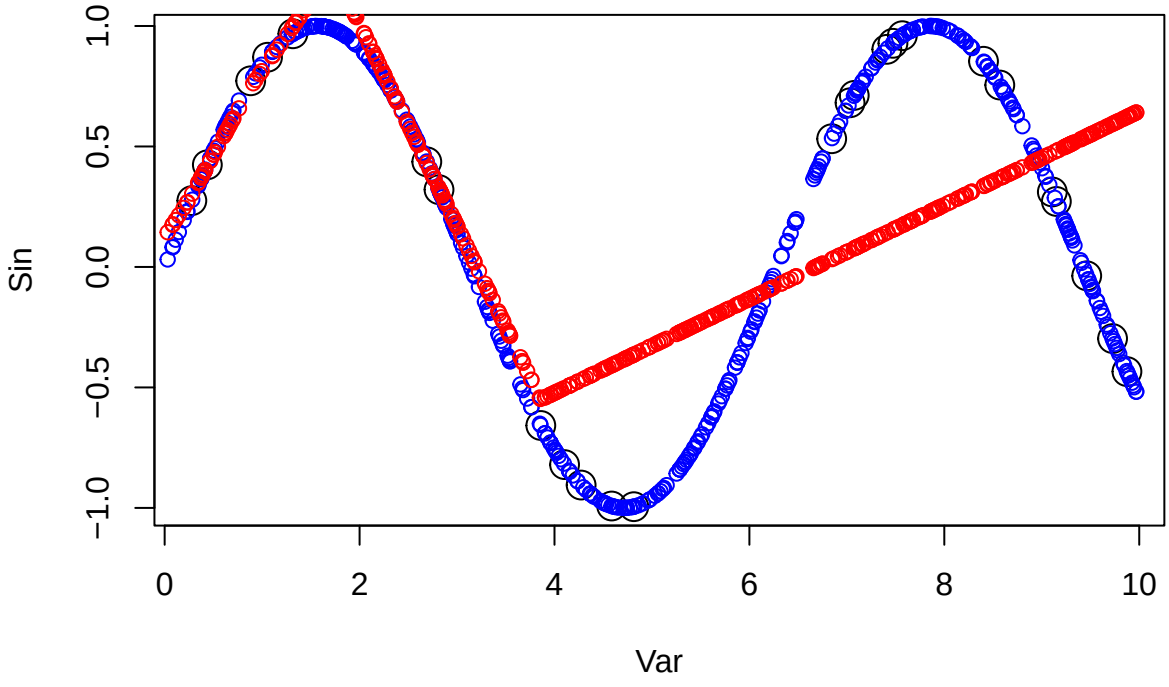
3.1 Task 1

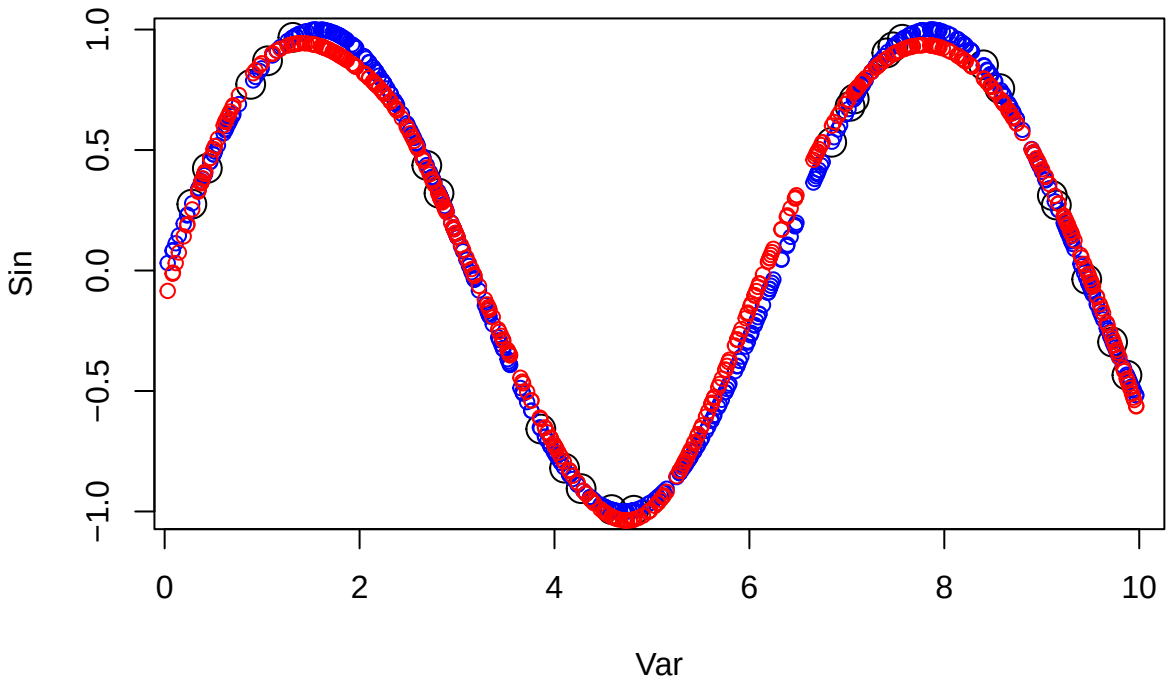


The model predictions are very good, except for a small deviation from var=5 to var=7, which is comparable to the sin values of the test set.

3.2 Task 2

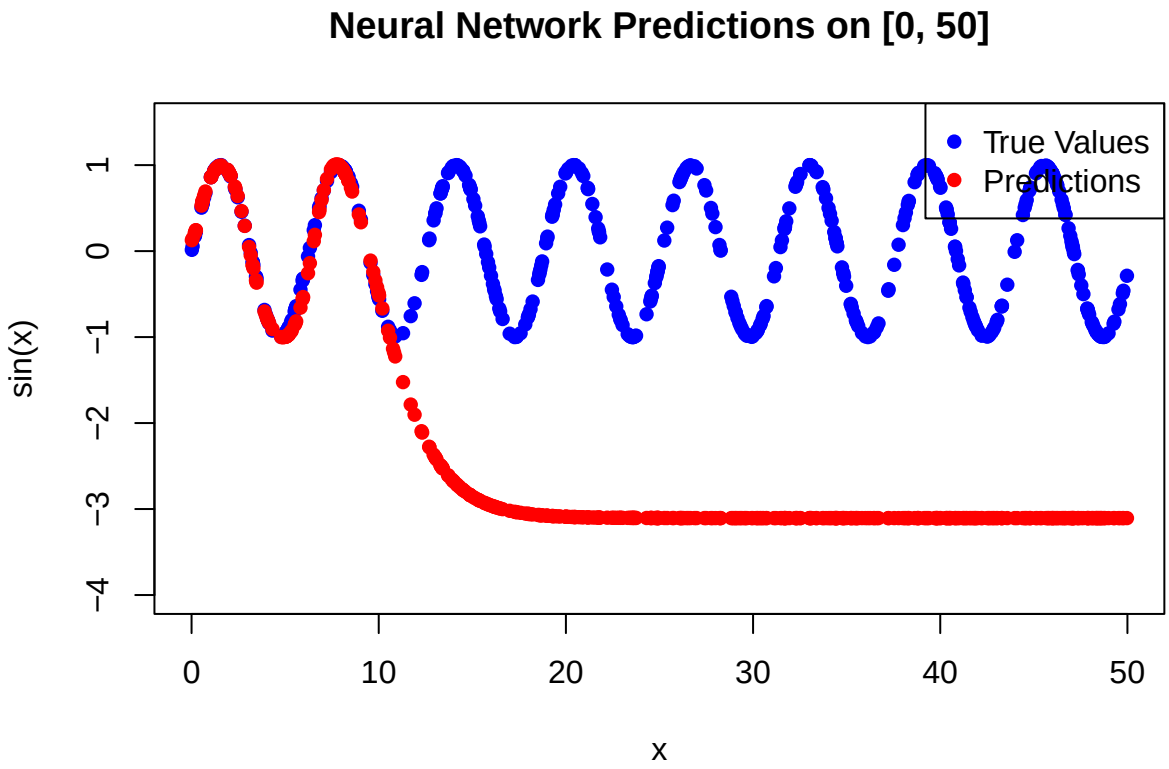






Linear function: The results of the linear function show that it cannot be used as an activation function at all, and the output of the neural network is not properly and appropriately expressed, resulting in a final output that is far from the labels of the test set and a distribution that does not match at all. ***Relu function:*** The Relu function performs well up to var=4, and after var=4, like the linear function, it is completely unusable as an activation function, i.e., it does not fit at all to the test set labeling distribution. ***Softplus function:*** The softplus function basically matches the curve of the test set labeling.

3.3 Task 3



Apparently, the model cannot predict based on the data points which are out of the scale [0,10], while it has good performance on the data points who fall in this interval although they have not appeared before.

3.4 Task 4: In question (3), the predictions seem to converge to some value. Explain why this happens.