Binary Classification

*Experiment Settings

Network: Single hidden layer network with ReLU activation function

Optimizer: Mini-Batch Stochastic Gradient Descent without momentum

Batch Size: 16

Learning Rate: Initial to 0.01

Learning Rate Scheduler: Exponential Decay with 0.95 each epoch

Epochs: 200

Loss Function: Cross Entropy Loss
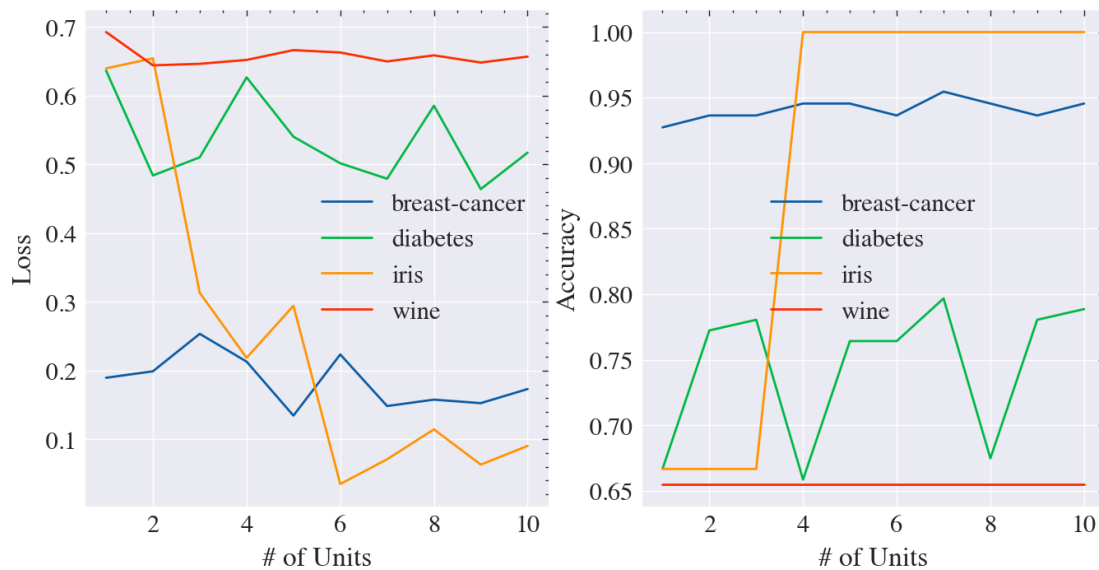
*Parameter Tuning Result

Figure 1 Loss & Accuracy on validation data

     Figure 1 demonstrates the development of the loss and accuracy of validation test split from training dataset according to each dataset and each number of hidden units. Parameters which lead to minimum loss value on the validation set will be chosen as the final parameters of the network. For "breast-cancer", H = 5 is chosen; for "diabetes", H = 9 is chosen; for "iris", H = 6 is chosen; for "wine", H = 2 is chosen. As shown as following table.

Table 1 # Hidden Layer chosen as final parameter

Dataset	# Hidden Layer chosen
Breast Cancer	5
Diabetes	9
Iris	6
Wine	2

*Performance on Test set

     Table 2 illustrates the performance on training set and test set respective to each dataset. And Figure 2 gives a straight-forward visualization for comparing conveniently.

Table 2 Performance on Training set and test set

Dataset	Train_loss	Train_acc	Test_loss	Test_acc

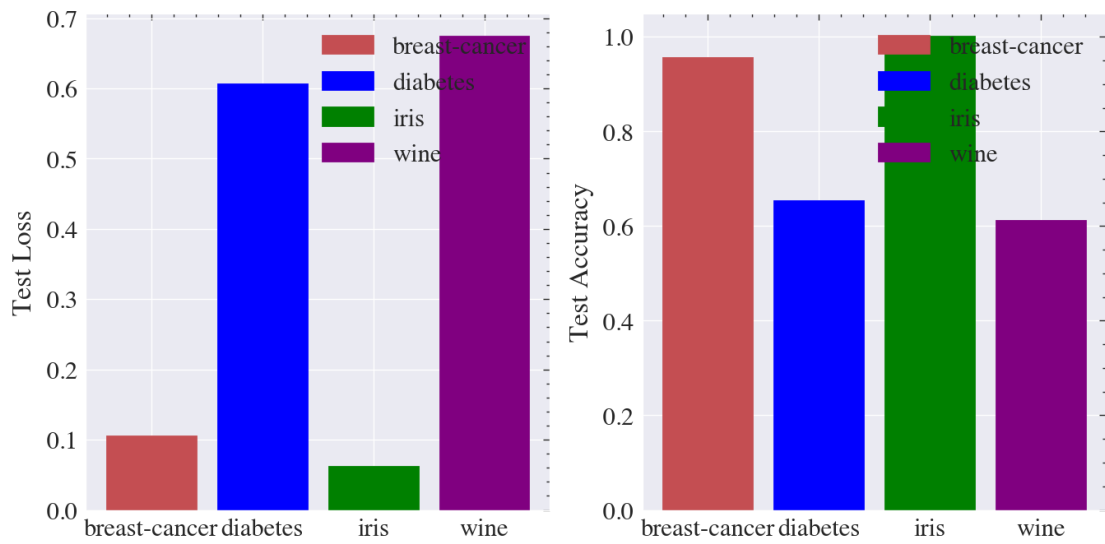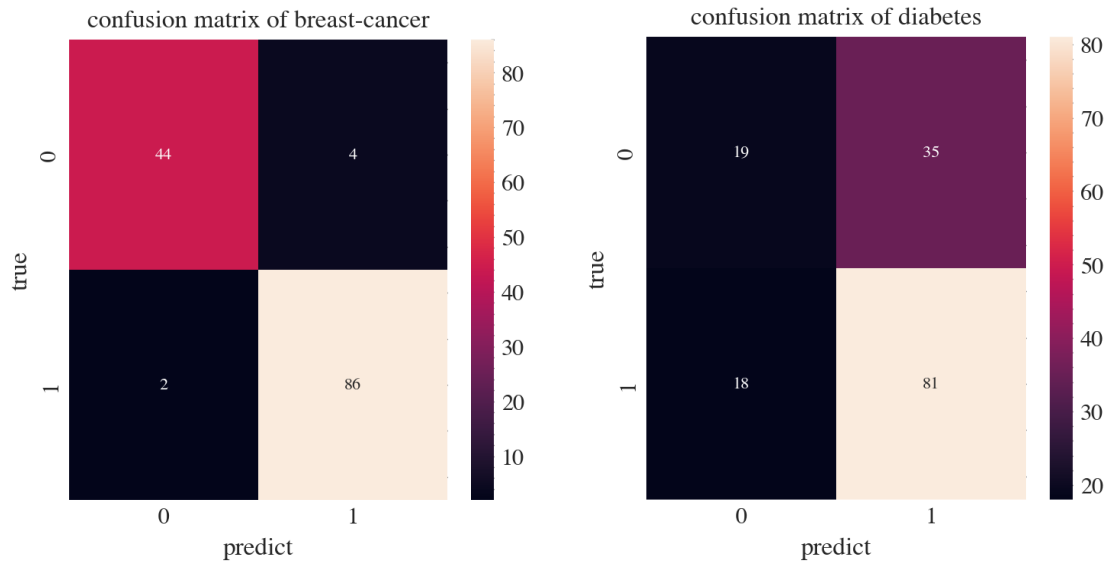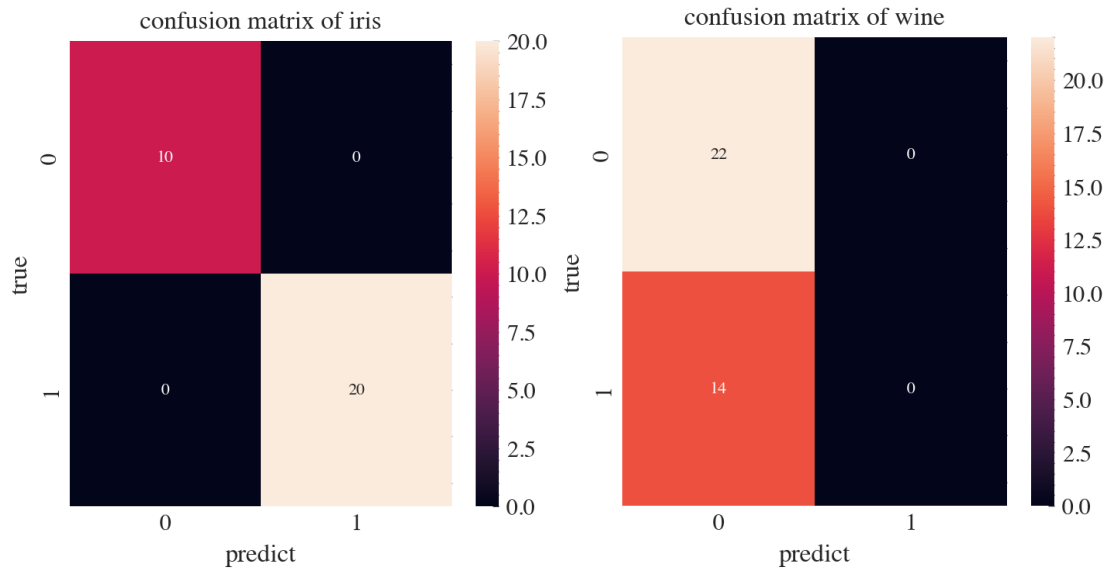|				
Breast Cancer	0.0984	0.9616	0.1054	0.9559
Diabetes	0.5480	0.7057	0.6057	0.6536
Iris	0.0476	1.0	0.0613	1.0
Wine	0.6765	0.5986	0.6734	0.6111

Figure 2 Visualization of model performance on each dataset

The following 4 figures demonstrate the confusion matrix of the prediction on test set, respectively.

confusion matrix of iris · confusion matrix of wine

Multi-class Classification

*Experiment Settings

Network: Double hidden layer network with ReLU activation function

Optimizer: Mini-Batch Stochastic Gradient Descent without momentum

Batch Size: 256

Learning Rate: Initial to 0.05

Learning Rate Scheduler: Exponential Decay with 0.98 each epoch

Epochs: 1000

Loss Function: Cross Entropy Loss

*Parameter Tuning Result

    Figure 3 demonstrates the loss & accuracy on the 20% validation data split from original training dataset. Same as the criterion mentioned in binary classification problem, combination leads to minimum loss value on validation dataset will be chosen as final parameters. The result suggests that the number of first hidden layer is 25 and the number of second hidden layer is 20.
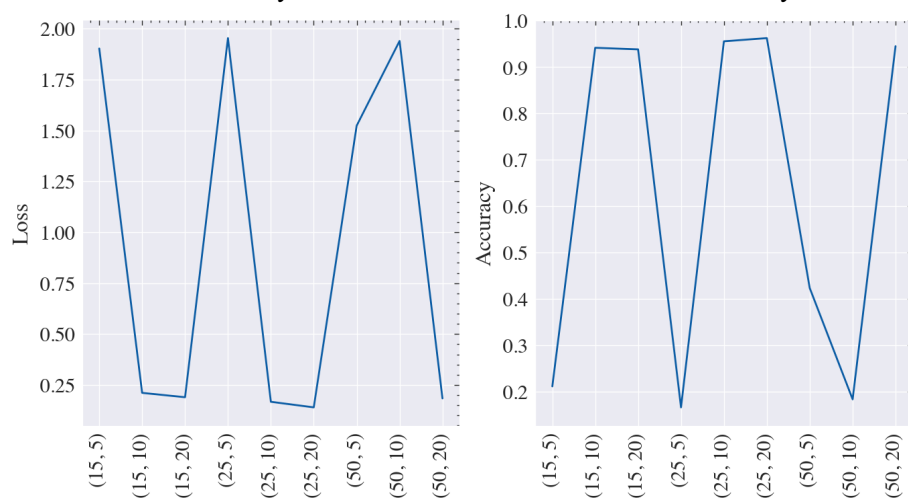

Figure 3 Loss & Accuracy on validation data respect to each combination of parameters

*Performance on Test set

Table 3 reports the performance of the final model on both training set and test set. Figure 4 illustrates the confusion matrix of the model on test set.

Table 3 Performance on Training set and test set

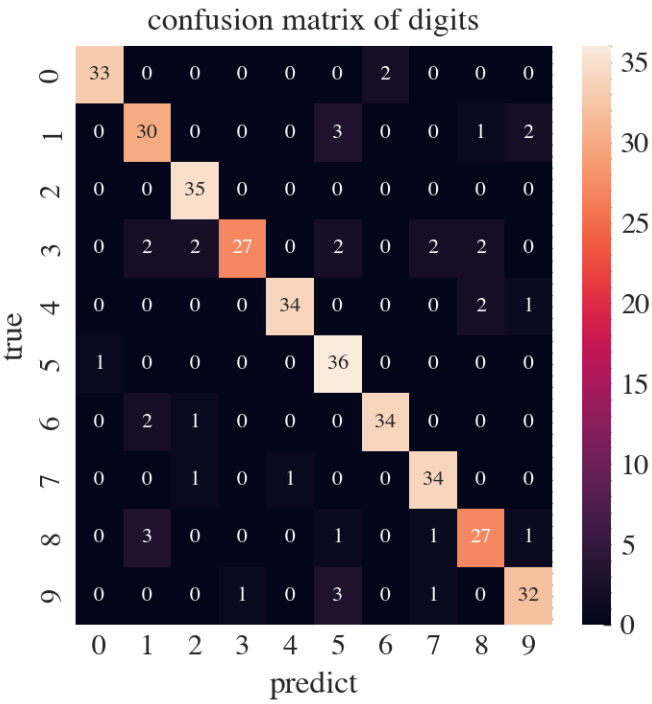| Dataset	Train_loss	Train_acc	Test_loss	Test_acc
Digits	0.0358	0.9930	0.4036	0.8944

Figure 4 Confusion matrix of digits