# Coding Assignment C3 Report

---

## 1.  Cross Validation Results

**1.1 Binary Classification:** The results of 5 fold cross validation on 5 binary classification datasets to find the best number of hidden units in the hidden layer (H) (Please refer to Notebook 1 in the code). The value of number of hidden units was varied from 1 to 10.

| Dataset | Best found number of hidden units | Average Cross Validation Accuracy across 5 folds |
|---|---|---|
| Iris | 9 | 95.83% |
| Digit | 9 | 96.25% |
| Breast Cancer | 10 | 96.71% |
| Wine | 7 | 73.12% |
| Diabetes | 8 | 77.07% |

**1.2 Multi class classification:** The results of 5 fold cross validation on the multi class classification datasets to find the best number of hidden units in layer first hidden layer (L1) and second hidden layer (L2) (Please refer to Notebook 3 in the code). The values of L1 were chosen from [50, 75, 100] and L2 is chosen from [10, 15, 20].

| Dataset | Best found number of hidden units in L1 | Best found number of Hidden units in L2 | Average Cross Validation Accuracy across 5 folds |
|---|---|---|---|
| Multi Class Dataset | 100 | 20 | 84.43% |

\* These are the results for best found from cross validation, for a full list of results please refer the notebooks in code.

---

## 2. Training Parameter Settings

After cross validation, during the actual training 2 configuration of hyper-parameters (learning rate, optimizer choice, number of epochs, batch size) were tried for all binary and multi-class classification datasets and chosen accordingly.

## 2.1 Binary Classification

For all 5 datasets, the same sets of configuration were used. For e.g., configuration 1 and 2 were defined once and used for all 5 datasets.
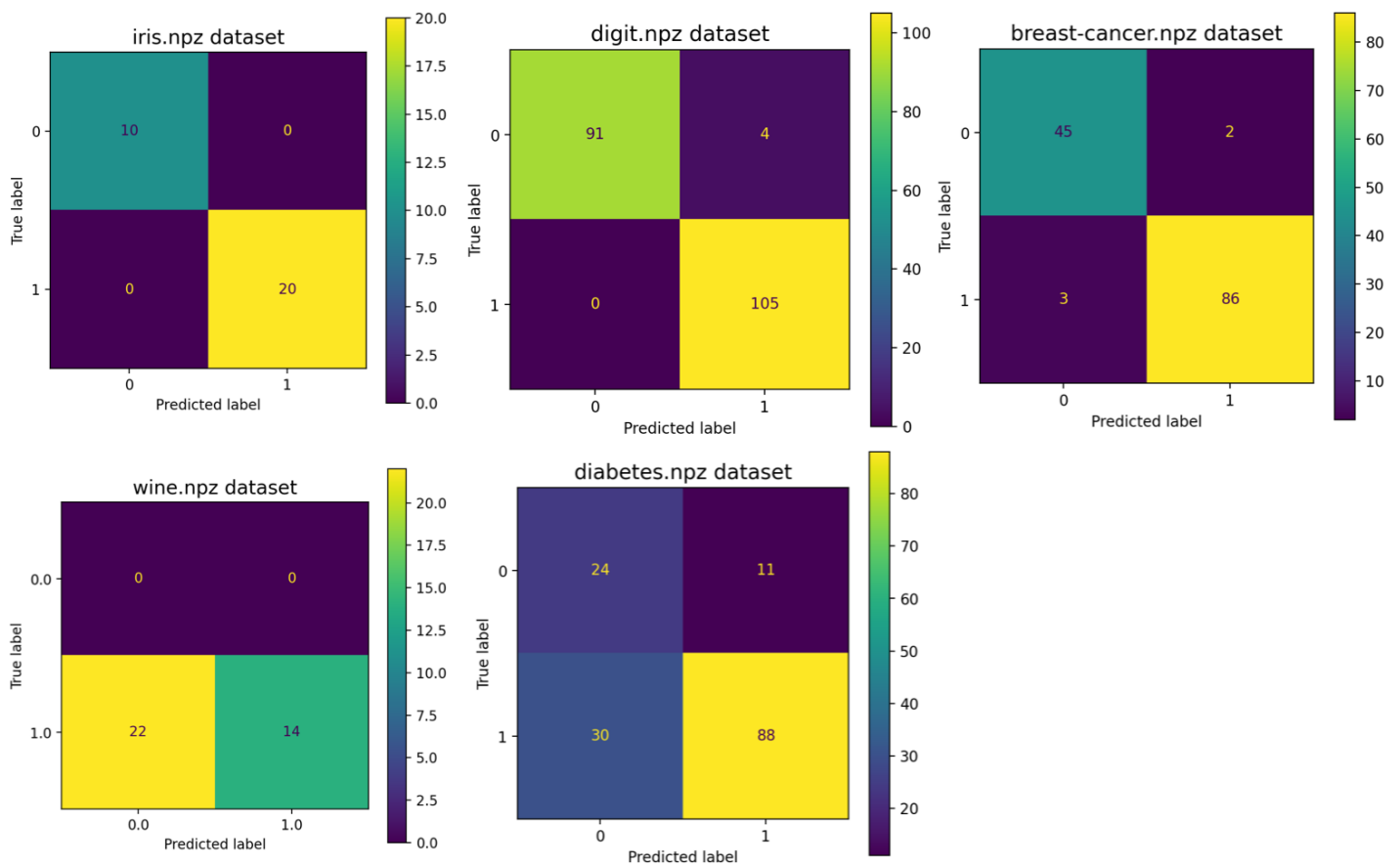
2.1.1. Training settings configuration 1

  Learning rate = 0.01
  Optimizer = Stochastic Gradient Descent (SGD)
  Number of Epochs = 30
  Batch size = 6

Configuration Matrices Obtained:
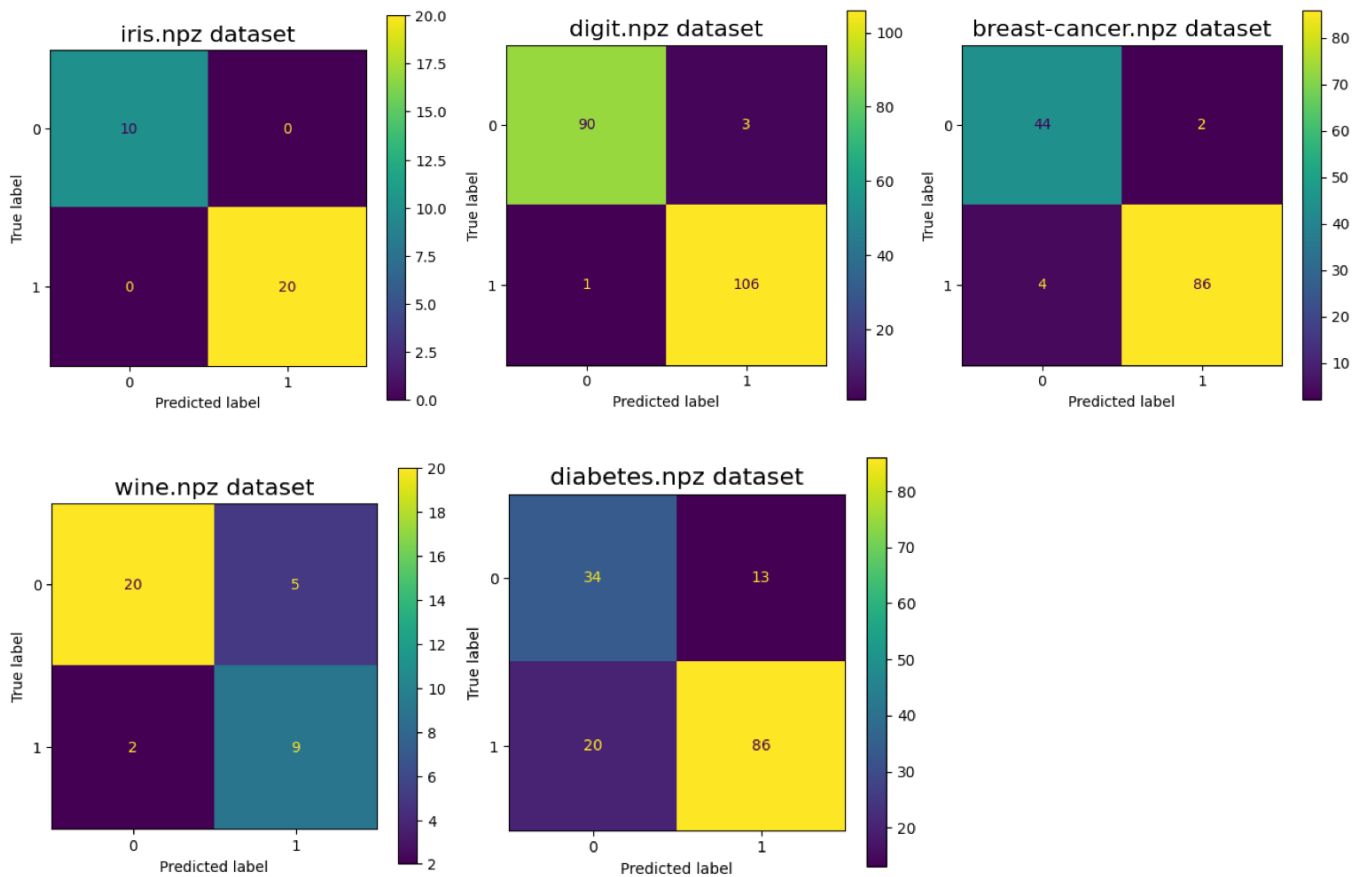


2.1.2 Training settings configuration 2

  Learning rate = 0.001
  Optimizer = Adaptive Moment Estimation (Adam)
  Number of Epochs = 20
  Batch size = 4

Confusion Matrices obtained:



By observing the confusion matrices for configuration 1 and 2, we can see that for iris, digit and breast-cancer datasets the performance is more or less the same in both configurations. But in configuration 1 the performance is significantly worse for wine and diabetes datasets than configuration 1. Hence, configuration 2 was chosen.

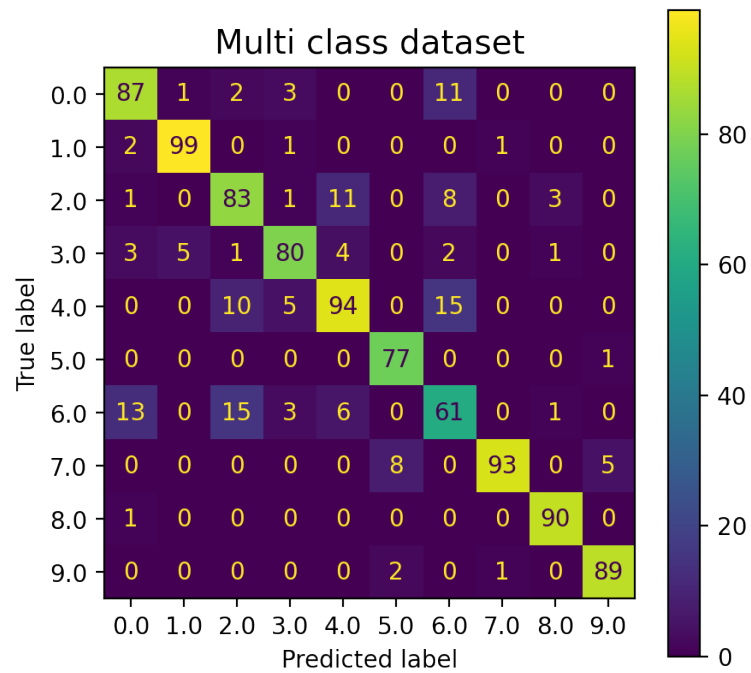**2.2 Multi Class Classification**
2.2.1 Training settings configuration 1
   Learning rate = 0.0001
   Optimizer = Adaptive Moment Estimation (Adam)
   Number of Epochs = 100
   Batch size = 32

Confusion matrix obtained:

## Multi class dataset

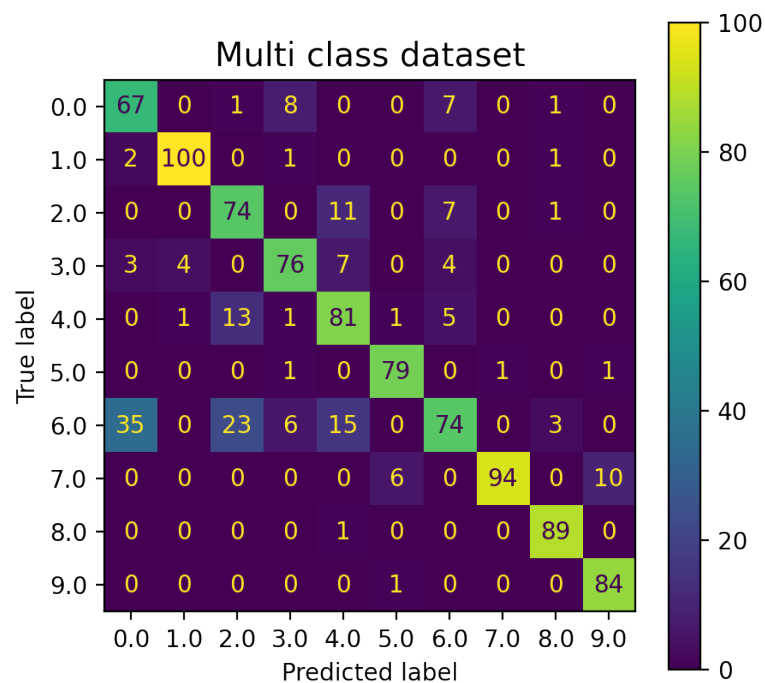|       | 0.0 | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | 7.0 | 8.0 | 9.0 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.0   | 87  | 1   | 2   | 3   | 0   | 0   | 11  | 0   | 0   | 0   |
| 1.0   | 2   | 99  | 0   | 1   | 0   | 0   | 0   | 1   | 0   | 0   |
| 2.0   | 1   | 0   | 83  | 1   | 11  | 0   | 8   | 0   | 3   | 0   |
| 3.0   | 3   | 5   | 1   | 80  | 4   | 0   | 2   | 0   | 1   | 0   |
| 4.0   | 0   | 0   | 10  | 5   | 94  | 0   | 15  | 0   | 0   | 0   |
| 5.0   | 0   | 0   | 0   | 0   | 0   | 77  | 0   | 0   | 0   | 1   |
| 6.0   | 13  | 0   | 15  | 3   | 6   | 0   | 61  | 0   | 1   | 0   |
| 7.0   | 0   | 0   | 0   | 0   | 0   | 8   | 0   | 93  | 0   | 5   |
| 8.0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 90  | 0   |
| 9.0   | 0   | 0   | 0   | 0   | 0   | 2   | 0   | 1   | 0   | 89  |

True label / Predicted label

2.2.2 Training settings configuration 2
Learning rate = 0.001
Optimizer = Adaptive Moment Estimation (Adam)
Number of Epochs = 70
Batch size = 16

Confusion matrix obtained:

## Multi class dataset

|       | 0.0 | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | 7.0 | 8.0 | 9.0 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.0   | 67  | 0   | 1   | 8   | 0   | 0   | 7   | 0   | 1   | 0   |
| 1.0   | 2   | 100 | 0   | 1   | 0   | 0   | 0   | 0   | 1   | 0   |
| 2.0   | 0   | 0   | 74  | 0   | 11  | 0   | 7   | 0   | 1   | 0   |
| 3.0   | 3   | 4   | 0   | 76  | 7   | 0   | 4   | 0   | 0   | 0   |
| 4.0   | 0   | 1   | 13  | 1   | 81  | 1   | 5   | 0   | 0   | 0   |
| 5.0   | 0   | 0   | 0   | 1   | 0   | 79  | 0   | 1   | 0   | 1   |
| 6.0   | 35  | 0   | 23  | 6   | 15  | 0   | 74  | 0   | 3   | 0   |
| 7.0   | 0   | 0   | 0   | 0   | 0   | 6   | 0   | 94  | 0   | 10  |
| 8.0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 89  | 0   |
| 9.0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 84  |

True label / Predicted label

Again, observing the confusion matrices obtained from configurations 1 and 2, we can see that configuration 1 is better than configuration 2. Hence configuration 1 was chosen.
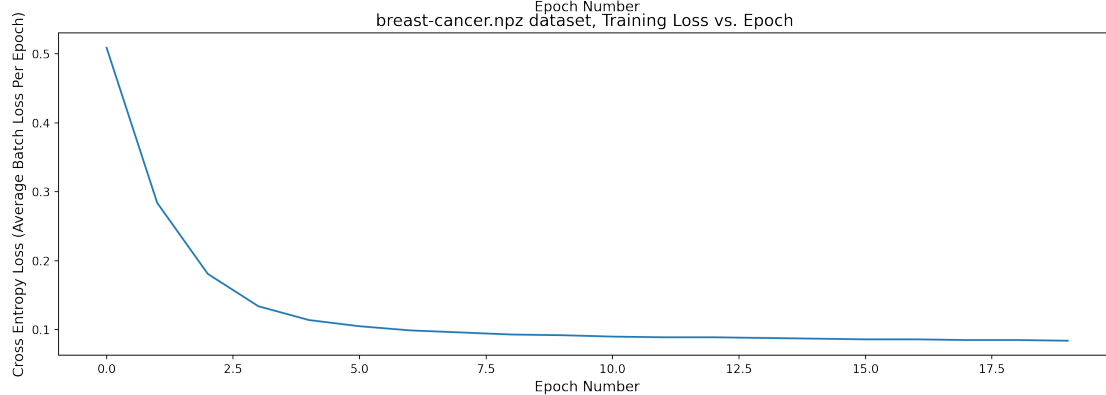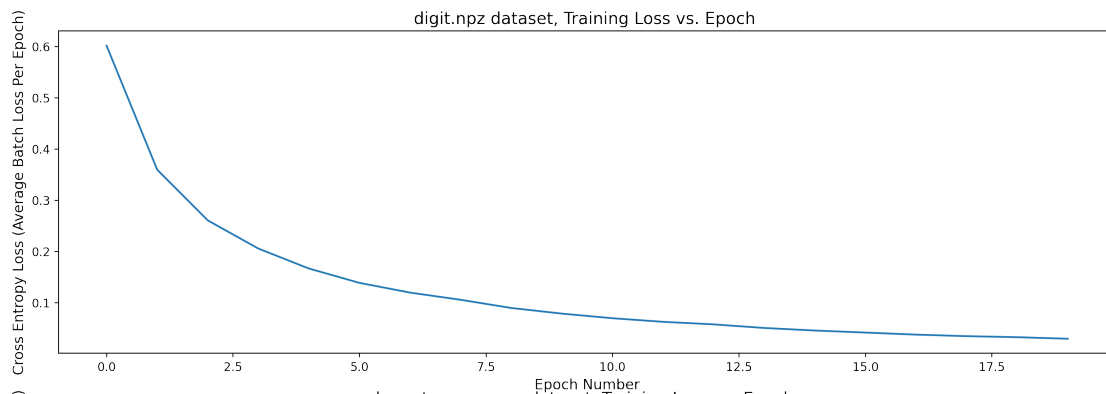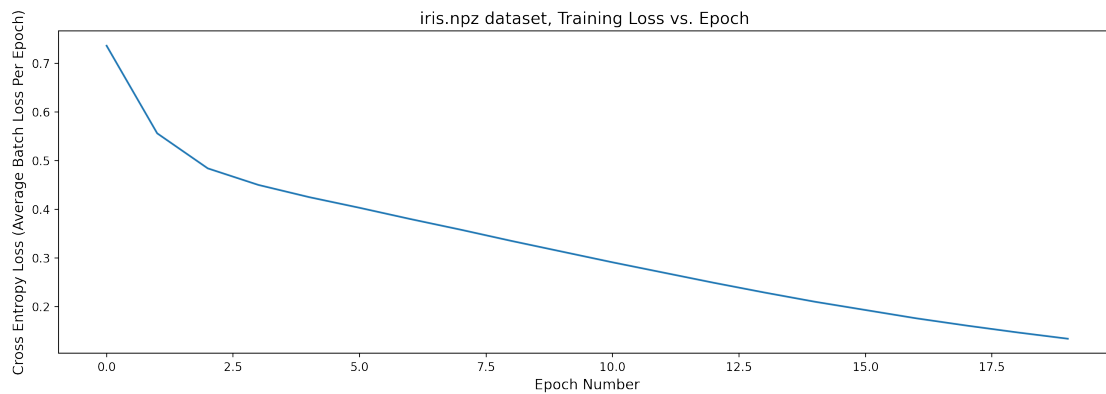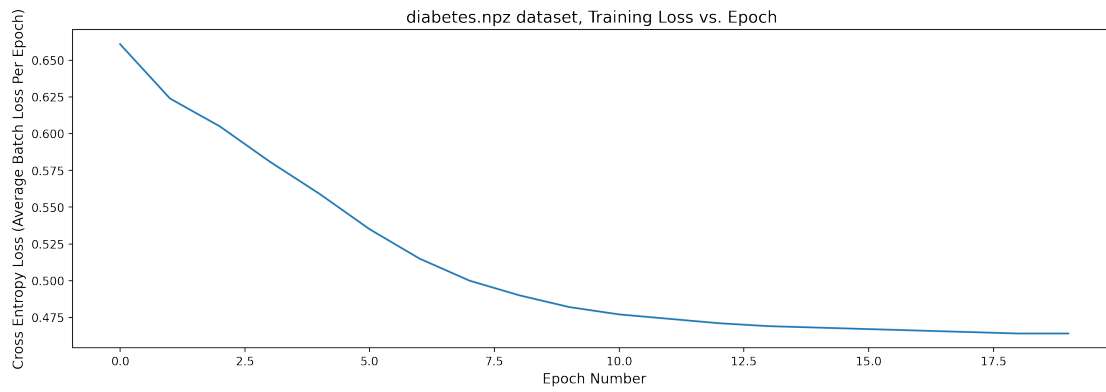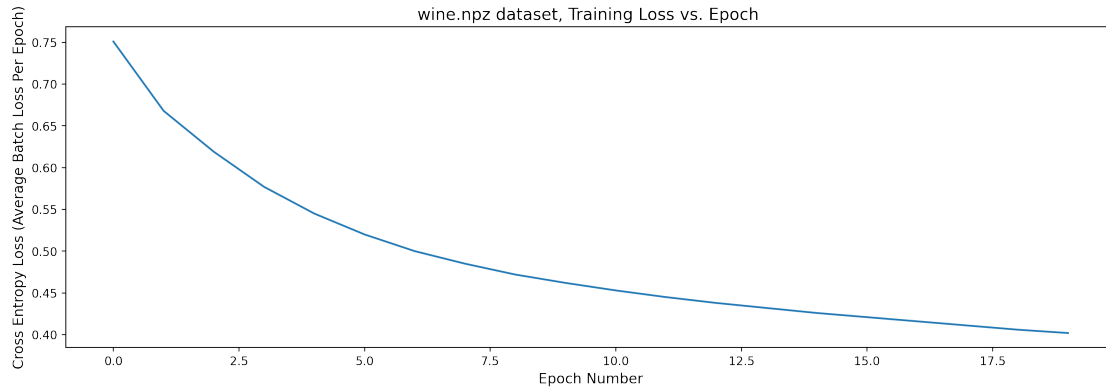
# 3. Training and test results

The results provided below in the 2 sections are based on the chosen configurations given in section 2.

**3.1 Cross Entropy results**
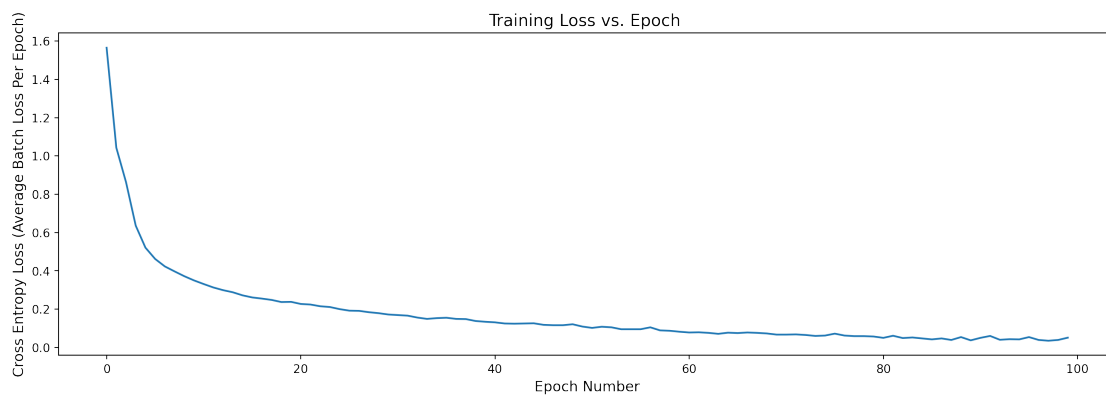
The cross entropy loss values were calculated during training and plotted in respective curves as given below:

3.1.1: Binary classification datasets

wine.npz dataset, Training Loss vs. Epoch



diabetes.npz dataset, Training Loss vs. Epoch

## 3.1.2 Multi class classification dataset



Training Loss vs. Epoch

In all the datasets, we can see that cross entropy is decreasing. This indicates that the respective MLP models were able to "learn" useful representations from data.

### 3.2 Accuracy results

3.2.1 Binary classification datasets

| Dataset | Training Accuracy | Test Accuracy |
|---|---|---|
| Iris | 100% | 100% |
| Digit | 99.37% | 98% |
| Breast Cancer | 97.07% | 95.58% |
| Wine | 82.39% | 80.55% |
| Diabetes | 76.58% | 78.43% |

1.2 Multi class classification dataset

| Dataset | Training Accuracy | Test Accuracy |
|---|---|---|
| Multi class dataset | 98.17% | 85.3% |

# 4. More Results

### 4.1 Training time
GPU: Not used
CPU type: Intel Core i7 10800K
CPU training times (in respective chosen configurations given in section 2):

| Dataset | Training Time (seconds) | Number of Epochs trained |
|---|---|---|
| Iris | 0.21 | 20 |
| Digit | 1.30 | 20 |
| Breast Cancer | 0.88 | 20 |
| Wine | 0.24 | 20 |
| Diabetes | 0.99 | 20 |
| Multi class dataset | 32 | 100 |

### 4.2 Early Stoping:
Early stopping was not used.