Assignment 1-Implementation of Backpropagation and Training a Palindrome Network

Jimut Bahan Pal, 22D1594
Shambhavi Pandey, 23D1145
Saikat Dutta, 23D2031

CS-772

20th February 2024

Problem Statement

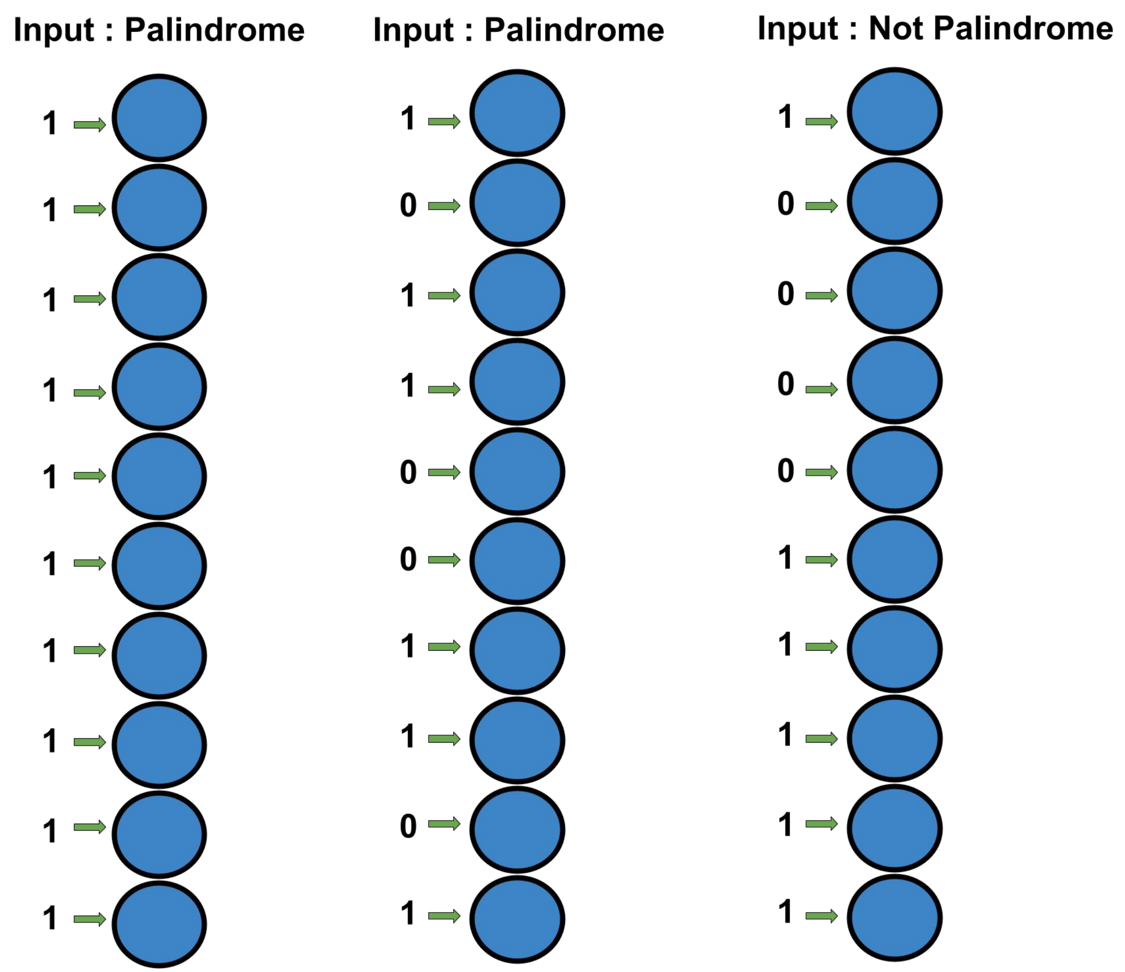
- Check if 10-bit string is palindrome or not
- **Technique used**: Neural network (BP from scratch)

Problem Statement

- **Input**: 10-bit String (of numbers)
- **Output**: Palindrome or not (class 0 or 1)
- There are about a total of 2^10 = 1024 input-output pairs:
  - Only 32 are having a class value of 1
  - 992 are having a class value of 0
  - Huge imbalance in the data!

Input representation

- The input is represented as a binary string of length 10.
- Each bit is fed to different nodes in the input layer.
- This is how the input is passed into the neural network



BP implementation

- Forward process:

$$z^1 = (w^1)^T x \tag{1}$$
$$a^1 = \phi(z^1) \tag{2}$$
$$z^2 = (w^2)^T a^1 \tag{3}$$
$$a^2 = \sigma(z^2) \tag{4}$$

Here, ϕ and σ are ReLU and Sigmoid activation respectively. w_{ij}^l connects i-th node in layer l to j-th node in layer $(l+1)$.

BP implementation

- Backpropagation:

Partial derivatives of loss w.r.t. weight parameters of layer 2 are given by,

$$\frac{\partial L}{\partial w_{i0}^2} = (a_0^2 - y_0)a_i^1 \tag{5}$$

Partial derivatives of loss w.r.t. weight parameters of layer 1 are given by,

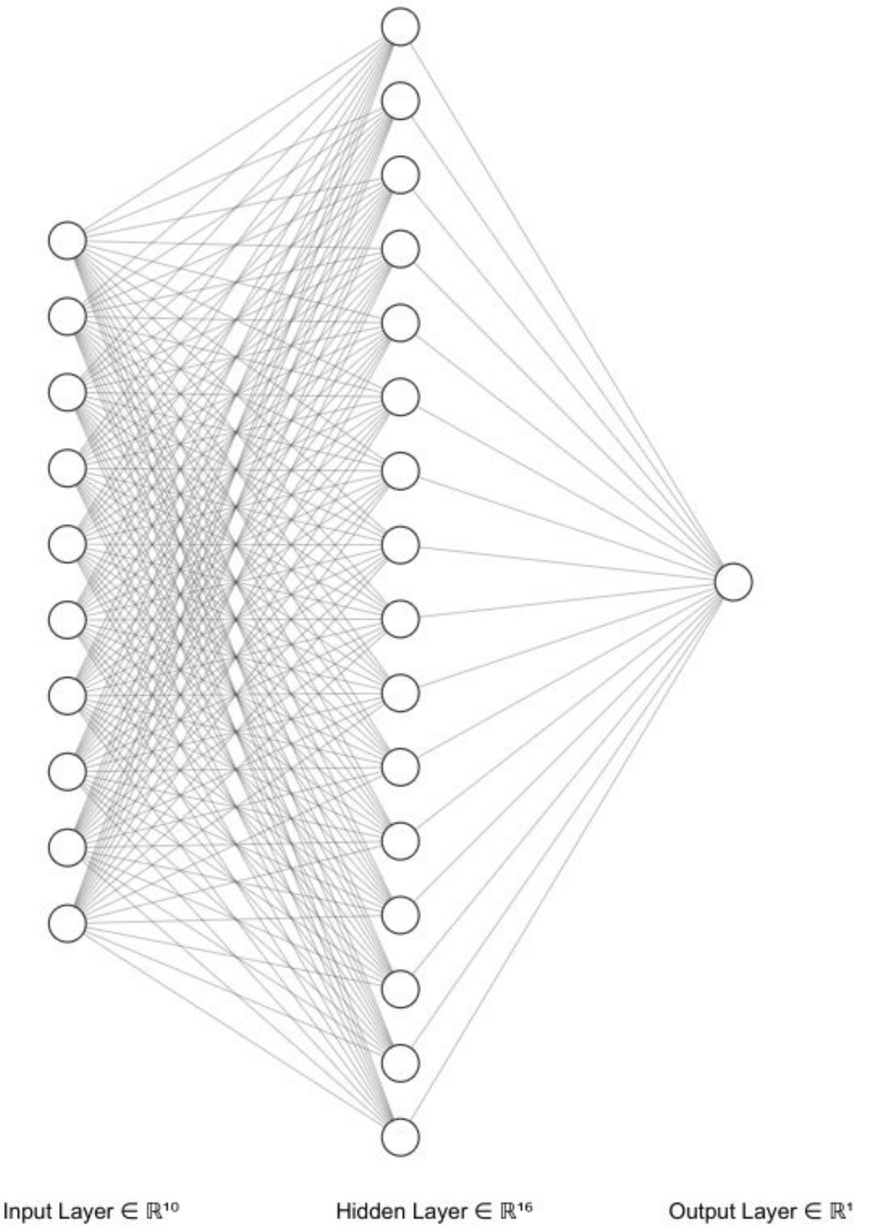
$$\frac{\partial L}{\partial w_{ij}^1} = (a_0^2 - y_0)w_{j0}^2\phi^{'}(z_j^1)x_i \tag{6}$$

BP implementation

- Other details:
  - Learning rate : 0.01
  - Momentum factor : 0.95
  - Epochs : 1000
  - Four-fold cross-validation is done
    - Positive-Negative sample ratio is kept same in each fold.

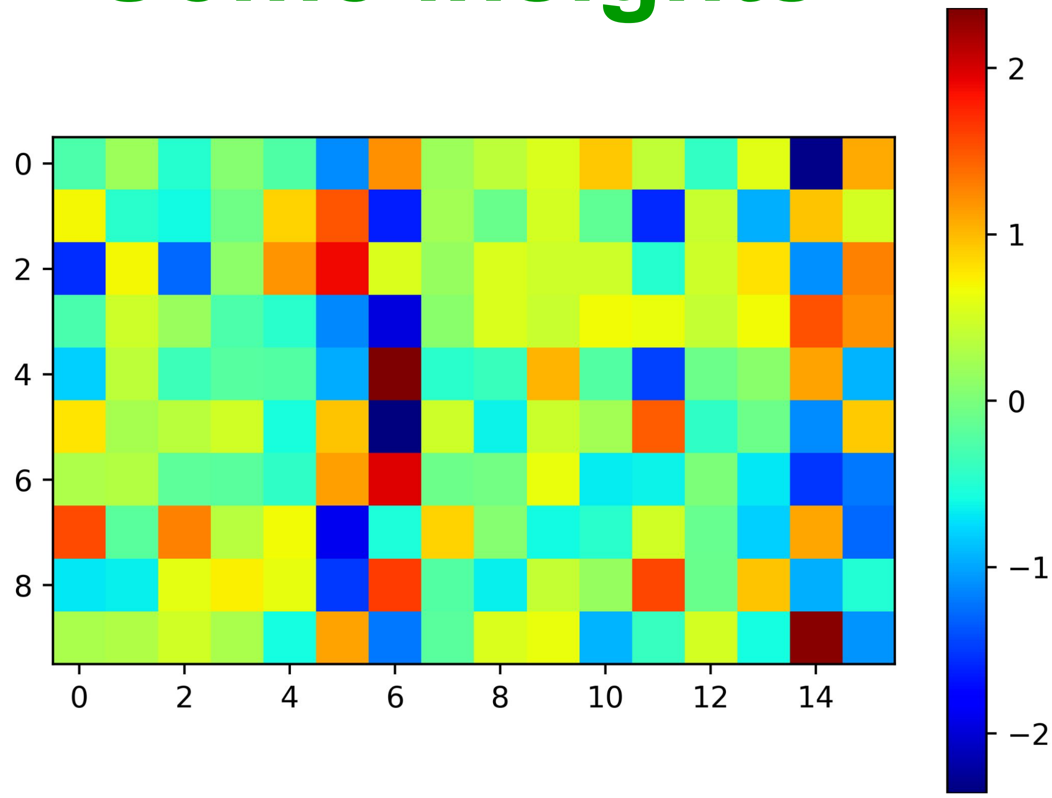
Architecture details

- Main architecture
  - Input layer: 10 neurons
  - Hidden layer : 16 neurons
  - Output layer : 1 neuron
  - ReLU is applied in hidden layer
  - Sigmoid is applied in output layer.

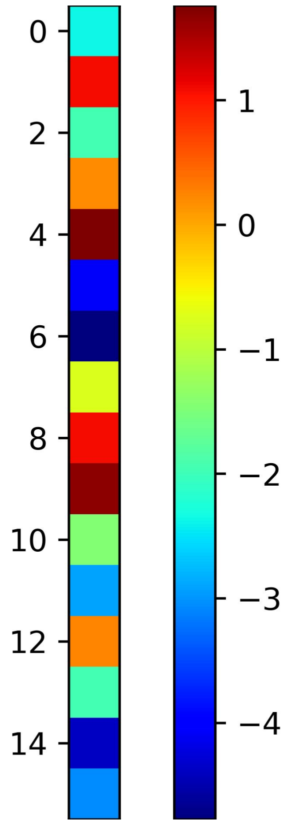


Some insights

- It is very hard to explain when there are 16 hidden neurons.



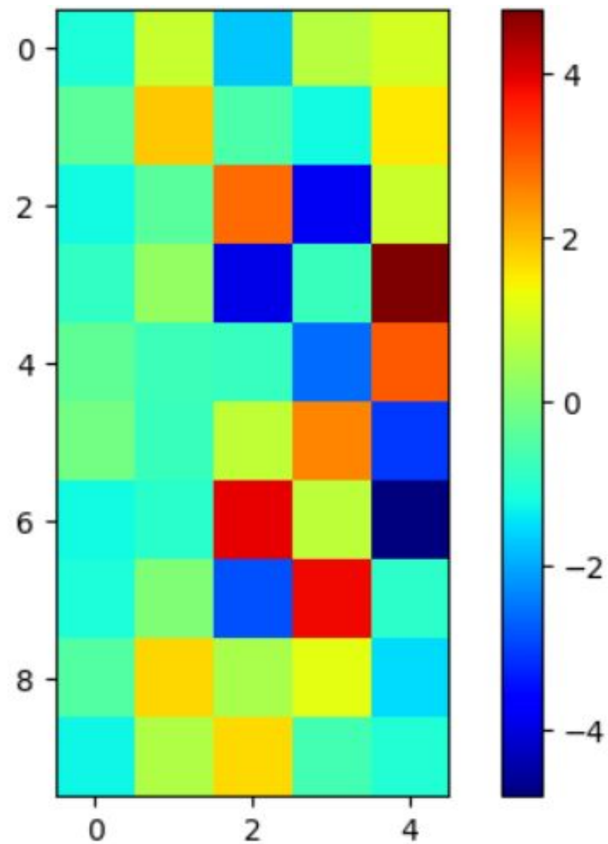
These are the weights connecting the **input to the hidden layer**



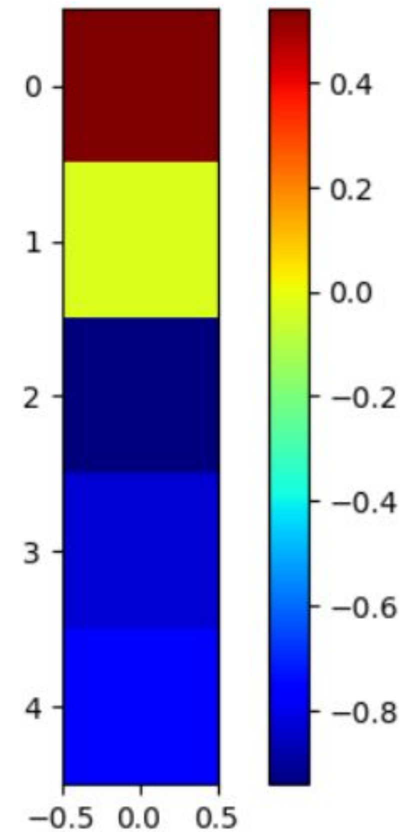
These are the weights connecting **the hidden to the output layer**

Some insights

- We find symmetry when **there are 5 hidden neurons.**
- Here we have initialized the weights between (-1,1).
- We have used bias term for 5 neuron neural network



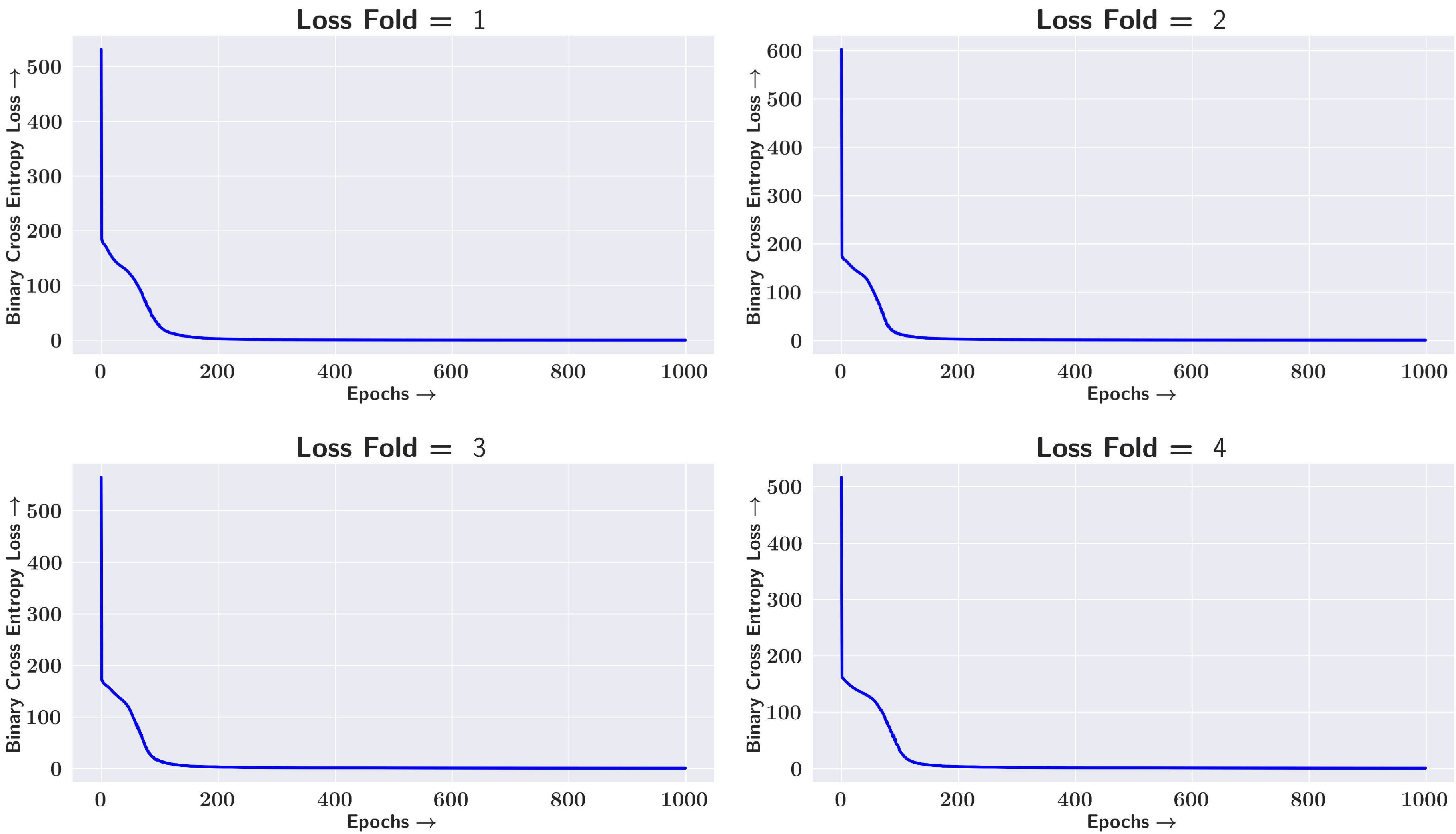
These are the weights connecting the **input to the hidden layer**



These are the weights connecting **the hidden to the output layer**

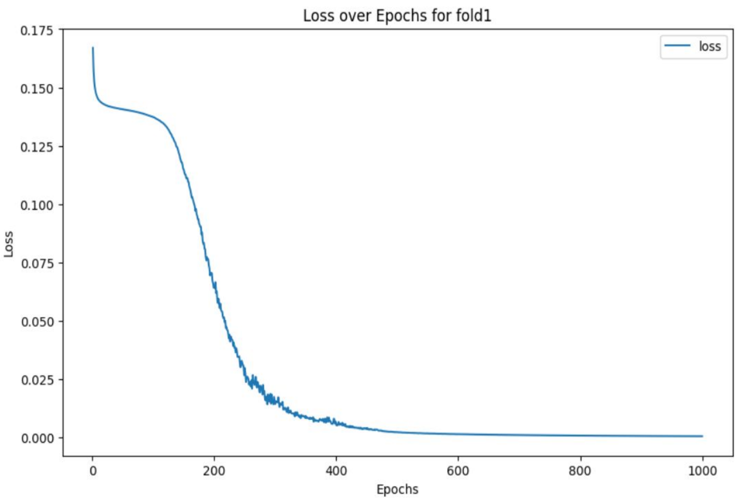
Some insights

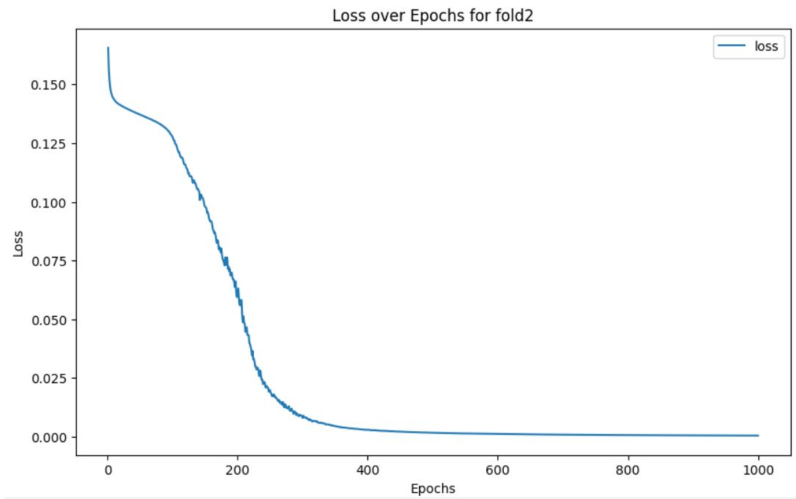
- Binary cross entropy loss is used and a 4-fold cross validation was used during training using 16 neurons

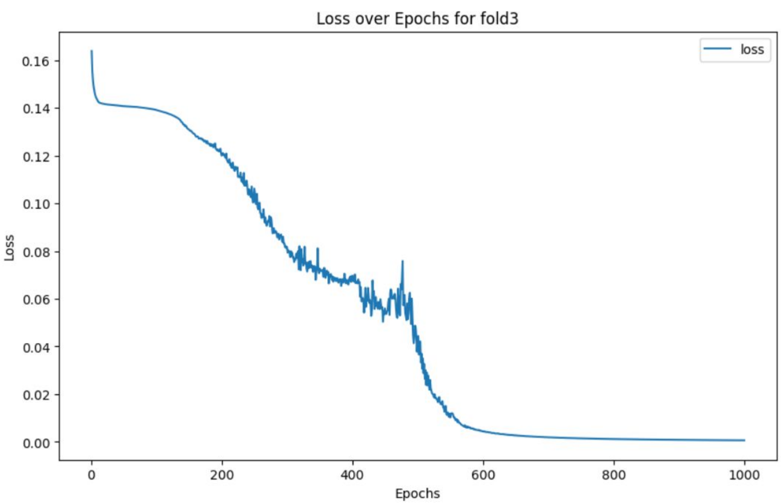


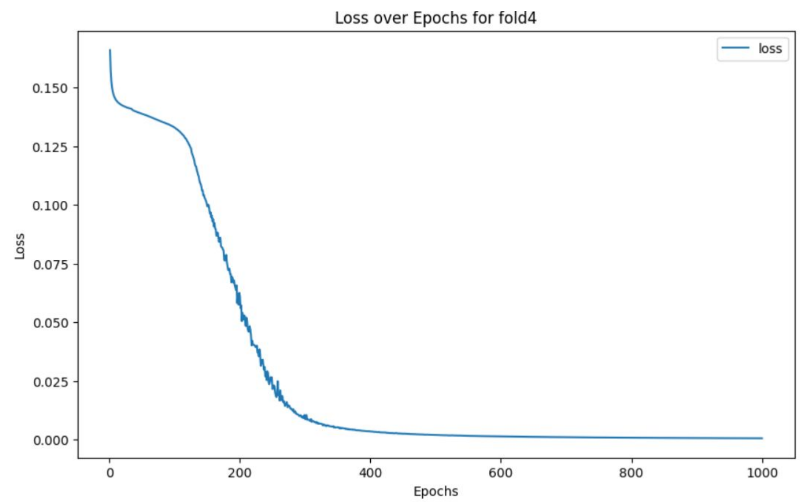
Some insights

- Binary cross entropy loss is used and a 4-fold cross validation was used during training using 5 hidden neurons.









Overall performance

- We have trained the model using 4-fold cross validation and on each fold we get a
  - Precision of 1.0
  - Recall of 1.0
  - F1-score of 1.0

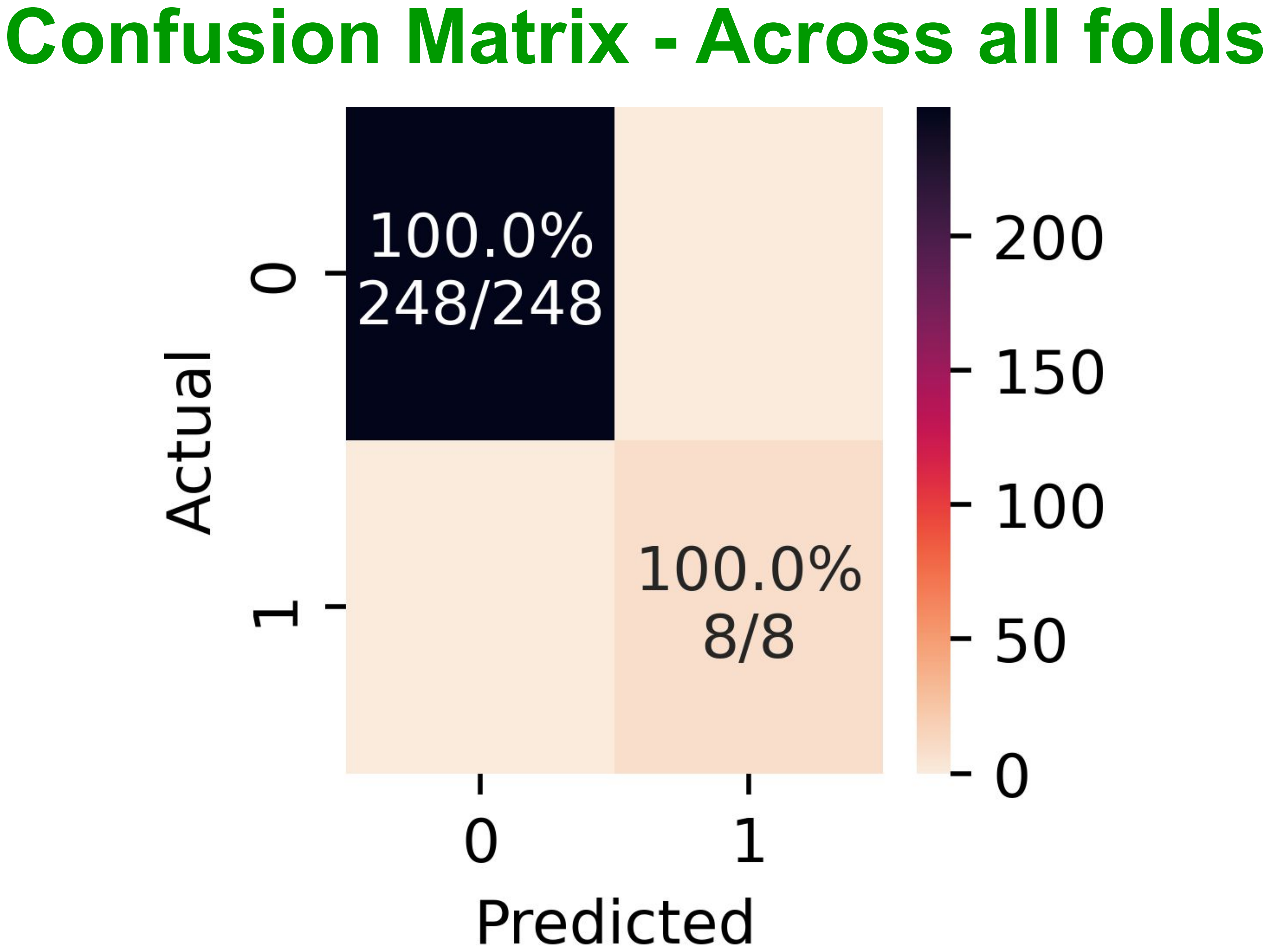
$$\text{Precision (PC)} = \frac{TP}{TP + FP}$$

$$\text{F1-Score} = \frac{2 \times (PC \times SE)}{PC + SE}$$

$$\text{Recall or Sensitivity (SE)} = \frac{TP}{TP + FN}$$

$$L_{y'}(y) := -\frac{1}{N}\sum_{i=1}^{N}(y'_i \log(y_i) + (1 - y'_i)\log(1 - y_i))$$

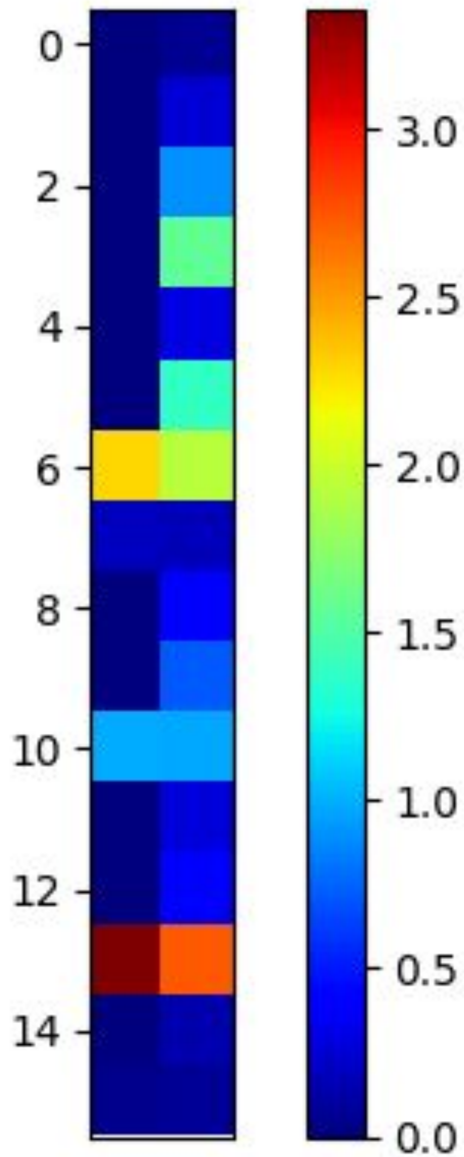
- Here, TP = True Positive, FP = False Positive.
- For Binary Cross Entropy loss function:
  - y’ is the ground truth (actual label) and
  - y is the predicted label (neural network’s output) of the class.



Confusion Matrix - Across all folds
100.0%
248/248
100.0%
8/8
Actual
0
1
Predicted
0
1
200
150
100
50
0

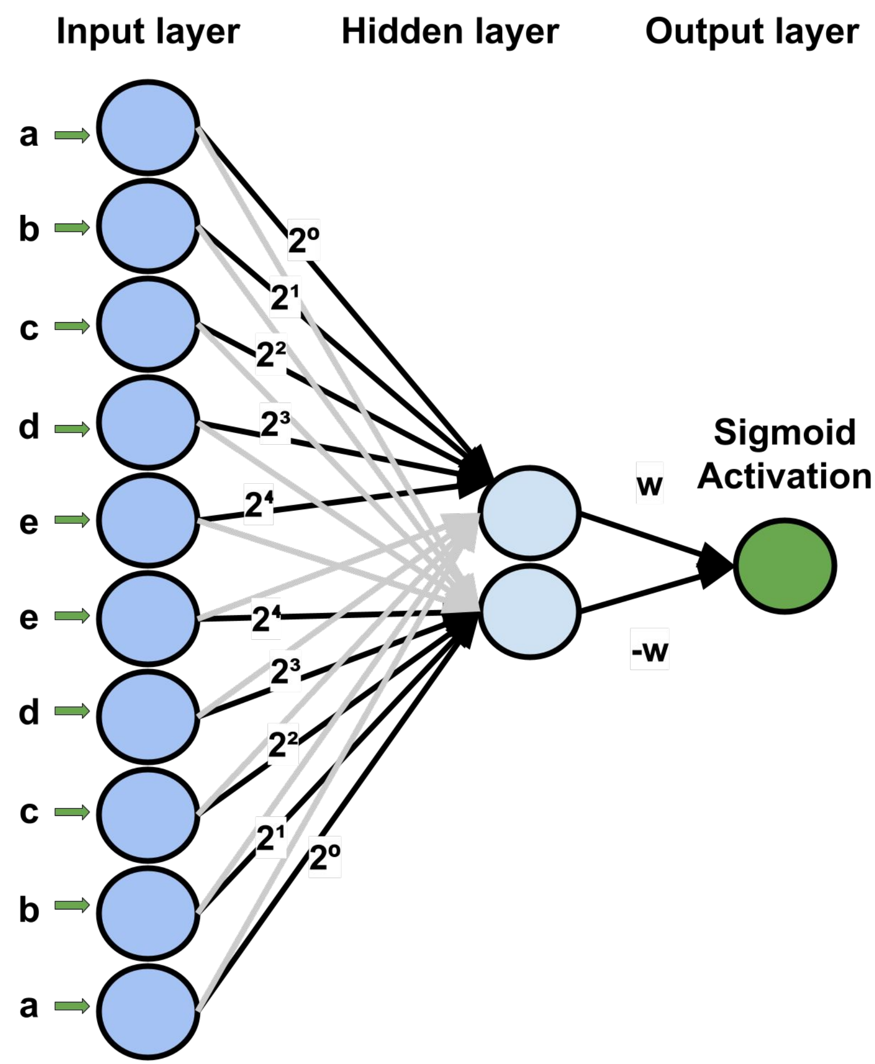
Interpretability of middle layer

- Left column: Average response for positive samples.
- Right column: Average response for negative samples.
- Only few nodes behave differently for positive and negative samples .



Learnings

- Theoretically it is possible using 2 neurons in hidden layer, but the weights are very hard to optimize.
- The palindrome number will have an unique symmetrical decimal representation
- We can use a **post processing technique** using delta to check if the output of the net is in between (0.5-δ,0.5+δ), then it is palindrome else it is not.



Thank You