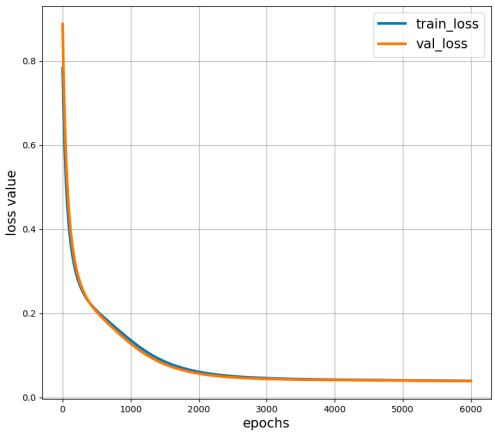
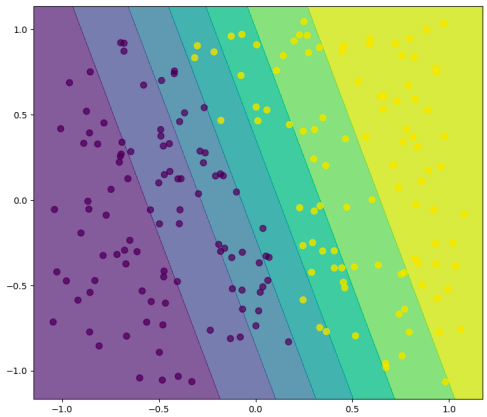
18786 – HW2 Report

Cheng An Hsieh(chengan2)

Deliverable 2.

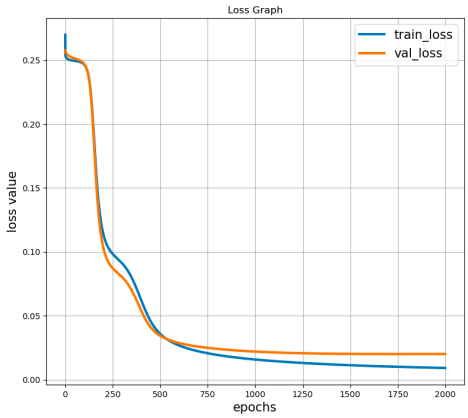
- MLP design and training details:
  - Optimizer: Basic Gradient Descent
  - Number of epochs: 6000
  - Learning rate: 1
  - Loss: l2
  - MLP design:
    - One hidden layer
    - Activation Function: Tanh
    - Hidden size: 1

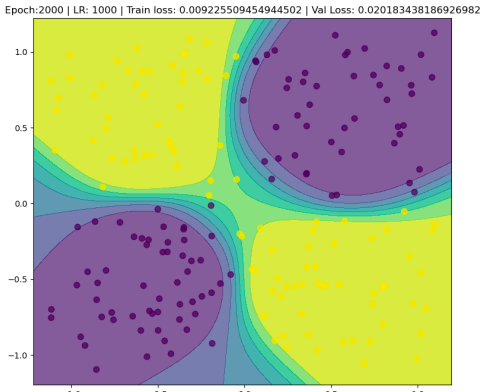




Deliverable 3.

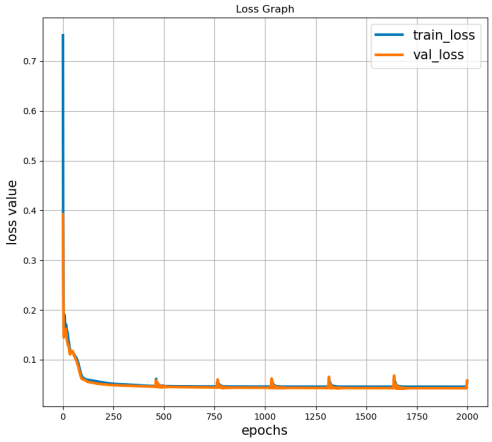
- MLP design and training details:
  - Optimizer: Basic Gradient Descent
  - Number of epochs: 2000
  - Learning rate: 1000
  - Loss: l2
  - MLP design:
    - One hidden layer
    - Activation Function: Sigmoid
    - Hidden size: 5

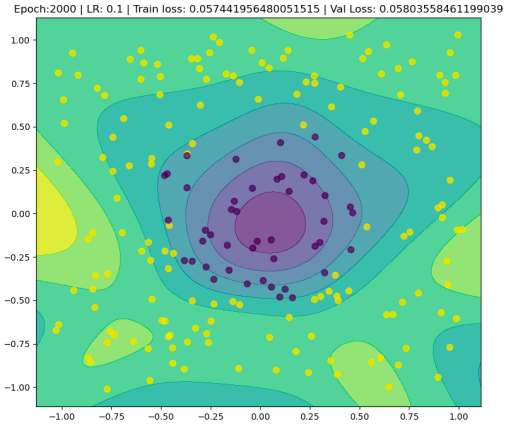




Deliverable 4.

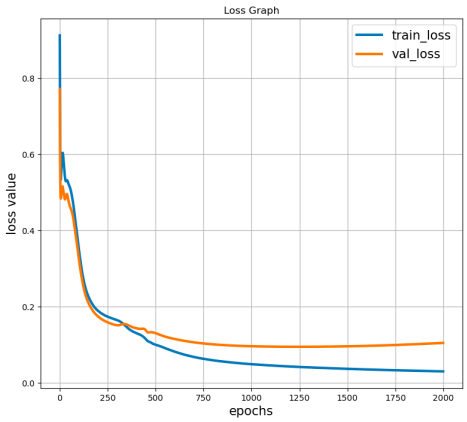
- **Regressor** - MLP design and training details:
  - Optimizer: Adam
  - Number of epochs: 2000
  - Learning rate: 1e-1
  - Loss: l2
  - MLP design:
    - One hidden layer
    - Activation Function: [tanh, linear]
    - Hidden size: 5

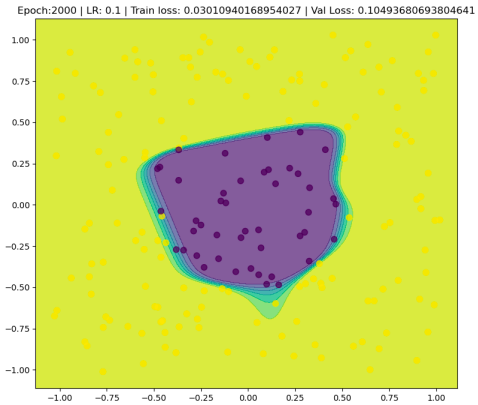




- **Classifier** - MLP design and training details:
  - Optimizer: Adam
  - Number of epochs: 2000
  - Learning rate: 1e-1
  - Loss: Cross-entropy
  - MLP design

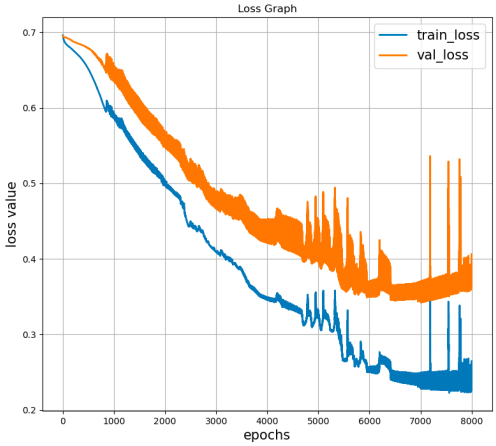
- One hidden layer
- Activation Function: [Sigmoid, Sigmoid]
- Hidden size: 5

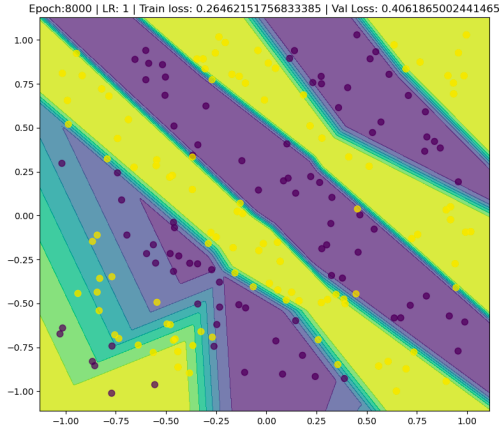




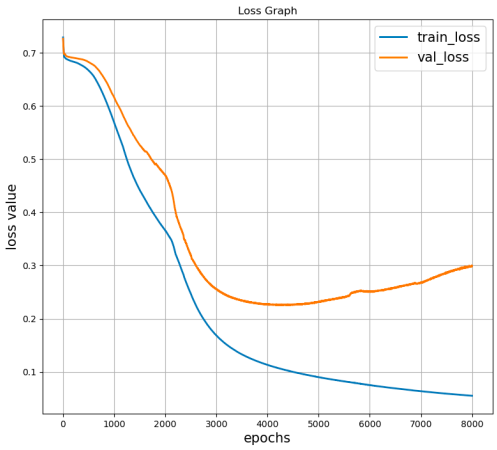
Deliverable 5.

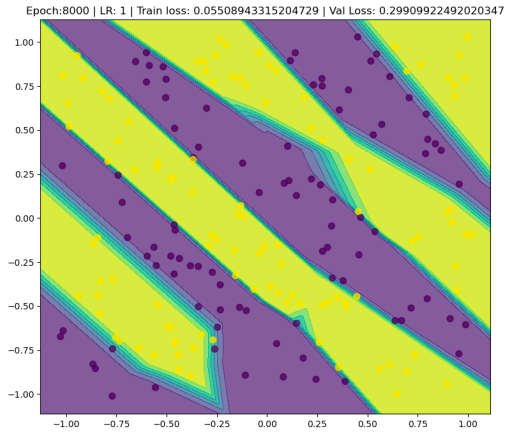
- Vanilla gradient descent - MLP design and training details:
  - Optimizer: vanilla gradient descent
  - Number of epochs: 8000
  - Learning rate: 1
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [2,6,4,1]
    - Activation Function: [relu, relu, sigmoid]



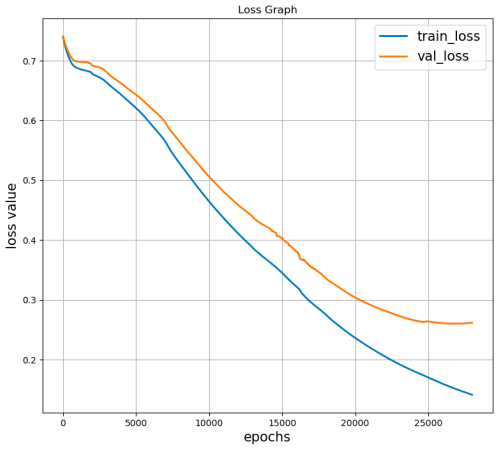


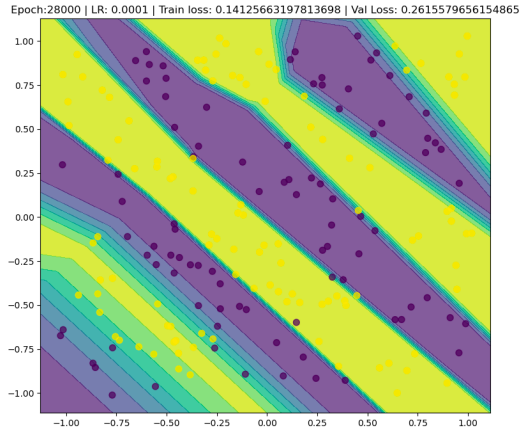
- Gradient descent with momentum - MLP design and training details:
  - Optimizer: Gradient descent with momentum
  - Number of epochs: 8000
  - Learning rate: 1
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [2,6,8,1]
    - Activation Function: [relu, relu, sigmoid]





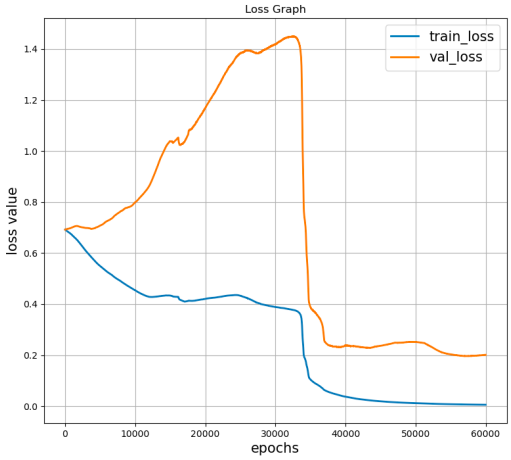
- Adam - MLP design and training details:
  - Optimizer: Adam
  - Number of epochs: 28000
  - Learning rate: 1e-4
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [2,10,8,1]
    - Activation Function: [relu, relu, sigmoid]

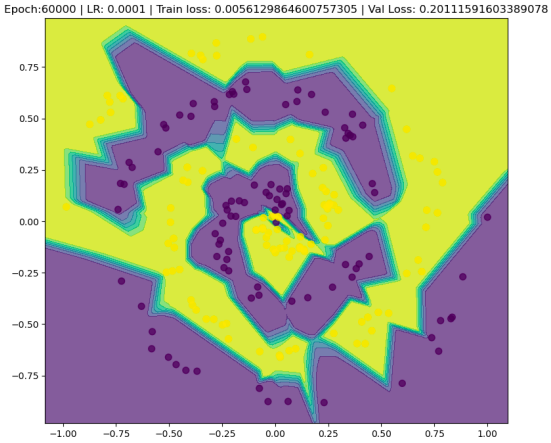




Deliverable 6.

- MLP design and training details:
  - Optimizer: Adam
  - Number of epochs: 60000
  - Learning rate: 1e-4
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [2,16,12,1]
    - Activation Function: [relu, relu, sigmoid]

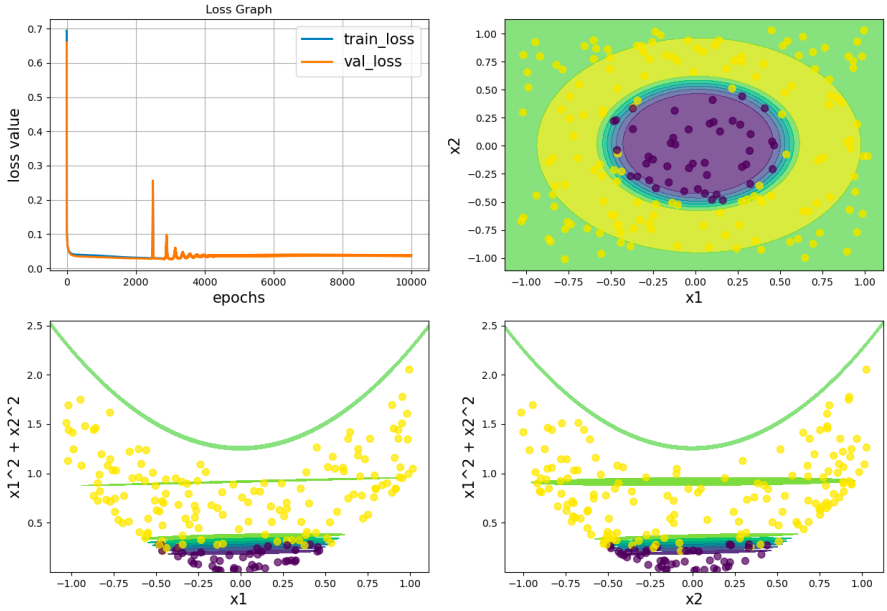




Deliverable 7.

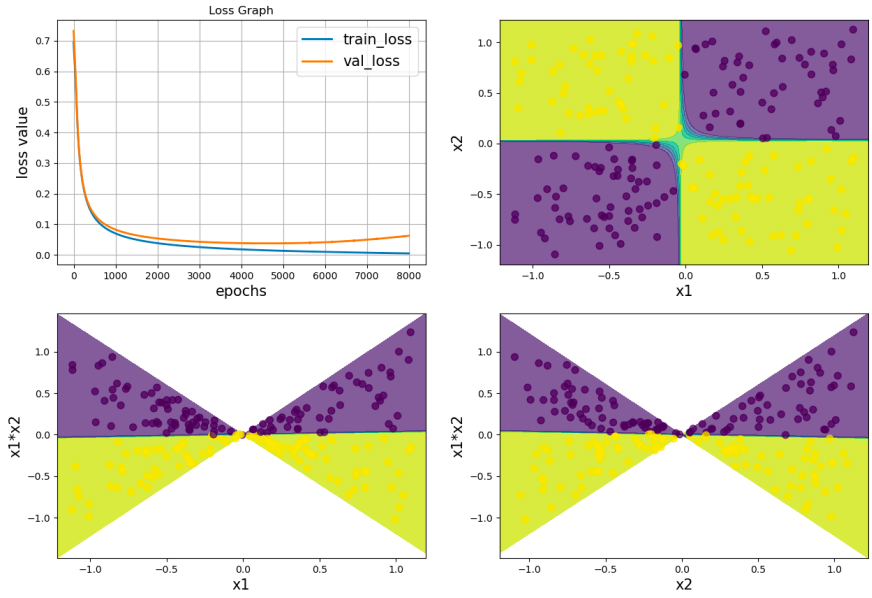
- Circles - MLP design and training details:
  - Nonlinear: x1^2 + x2^2
  - Optimizer: Basic Gradient Descent
  - Number of epochs: 10000
  - Learning rate: 100
  - Loss: l2
  - MLP design:
    - Two hidden layers: [3, 2, 1]
    - Activation Function: [tanh, linear]

Epoch:10000 | LR: 100 | Train loss: 0.03650128149717199 | Val Loss: 0.03867029217774451



- XOR - MLP design and training details:
  - Nonlinear: x1 * x2
  - Optimizer: Adam
  - Number of epochs: 8000
  - Learning rate: 1e-2
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [3, 1, 1]
    - Activation Function: [linear, sigmoid]

Epoch:8000 | LR: 0.01 | Train loss: 0.0049119013581315416 | Val Loss: 0.0626915529621581



- Swiss-roll - MLP design and training details:
  - Nonlinear: sqrt (x1^2 + x2^2) * sign(x1)
  - Optimizer: Adam
  - Number of epochs: 41000
  - Learning rate: 1e-4
  - Loss: cross-entropy
  - MLP design:
    - Two hidden layers: [3, 10, 8, 1]
    - Activation Function: [relu, relu, sigmoid]