

## Neural Networks - Project Answers

### I. True or False

1. **True**
  2. **False**
  3. **True**
  4. **False**
  5. **False**
  6. **True**
  7. **True**
  8. **False**
  9. **True**
  10. **True**
- 

### II. Match the options

Match each phenomenon with a plot:

1. **a (1-iv, 2-ii, 3-i, 4-iii)**
  2. **c (1-linear, 2-sigmoid, 3-tanh, 4-ReLU, 5-softplus)**
  3. **c. (f = activation, W = weight/parameter, x = input, b = bias)**
- 

### III. Fill in the Blanks

1. **Data augmentation**
  2. **Generator, discriminator**
  3. **Forget gate**
  4. **Loss function**
  5. **Backpropagation**
  6. **Sequential**
  7. **Weight decay (L2 regularization)**
  8. **ReLU**
- 

### IV. Short Answer Questions

1. **L1 Regularization vs. L2 Regularization:**
  - L1 regularization (Lasso) is preferred for feature selection, as it drives irrelevant weights to zero.

- L2 (Ridge) is better for preventing overfitting by discouraging large weights without inducing sparsity.
- 2. **Reason for a neural network fail to approximate a function**
  - A neural network may fail due to insufficient model capacity, poor optimization like bad initialization.
- 3. **Difference between learning and pure optimization:**
  - Learning differs from optimization because learning involves generalization to training data, while optimization only minimizes the loss on the training set.
- 4. **saturation of sigmoid not a concern in the output layer**
  - In the output layer, sigmoid is typically used for binary classification, where saturation is mitigated in the output layer because loss functions like cross-entropy directly adjust gradients.
- 5. **Number of random numbers for dropout:**
  - If there are  $n$  layers, each with  $m$  units, and you take  $s$  gradient steps, then the number of random dropout masks needed is  $\mathbf{n \times m \times s}$ .