Neural Networks - Project

[Marks – 20]

I. True or False: [Marks = 0.5 X 10 = 5]

1. Empirically, deeper networks perform better than wider networks.
- True
- False

2. Parameter norm penalties apply hard constraints on parameter values.
- True
- False

3. The main reason to use back-propagation is to avoid repeated calculations.
- True
- False

4. Local minima are a major challenge for gradient descent in neural network training.
- True
- False

5. Mini-batch SGD with a stream of training data minimizes generalization error.
- True
- False

6. Dropout typically requires more iterations of the training algorithm.
- True
- False

7. Momentum can accelerate learning in the face of high curvature.
- True
- False

8. Gradient clipping is the solution to saddles in the cost function.
- True
- False

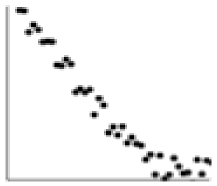
9. The default recommendation for an activation function is ReLU.
- True
- False

10. A deep feedforward network represents the composition of many simple functions.
- True
- False

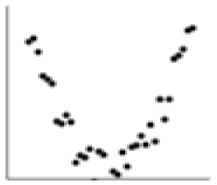
II. Match the options with correct picture and approach: [Marks = 2 X 3 = 6]

1. Match each phenomenon with a plot. Each plot shows the norm of the gradient (on the y axis) over the course of training a neural network model.
    i. encountering an ill conditioned Hessian
    ii. passing a saddle point
    iii. passing a cliff
    iv. approaching a local minimum

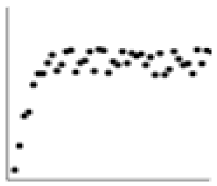
1.



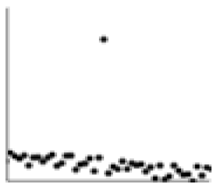
2.



3.



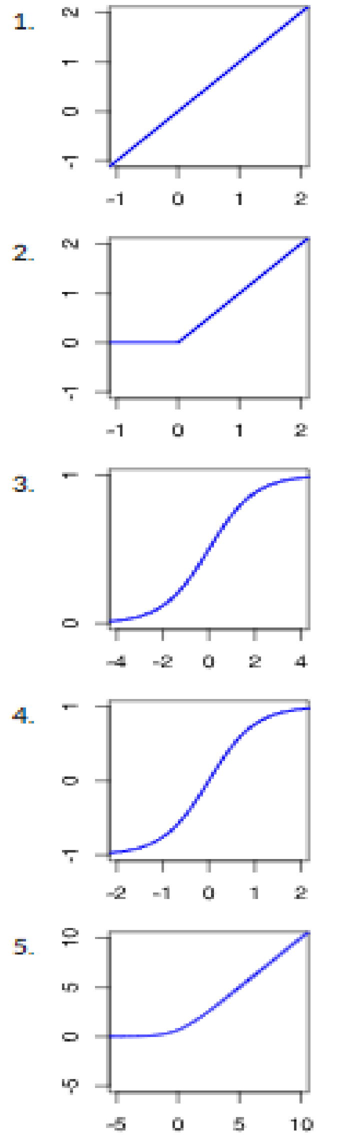
4.



Options:

a. 1-iv, 2-ii, 3-i, 4-iii
b. 1-i, 2-ii, 3-iii, 4-iv
c. 1-iv, 2-iii, 3-i, 4-ii
d. 1-i, 2-iii, 3-iv, 4-ii

2. Match each activation function with its plot
    i. ReLU
    ii. tanh
    iii. linear
    iv. sigmoid
    v. softplus



Options:

a. 1-ReLU, 2-tanh, 3-sigmoid, 4-softplus, 5-linear
b. 1-linear, 2-ReLU, 3-sigmoid, 4-tanh, 5-softplus
c. 1-linear, 2-sigmoid, 3-tanh, 4-ReLU, 5-softplus
d. 1-ReLU, 2-linear, 3-sigmoid, 4-softplus, 5-tanh

3. Below is the equation for part of the neural network

$$\mathbf{o} = f(\mathbf{W}^{\top}\mathbf{x} + \mathbf{b})$$

Give the term for the specified part of the equation

i. activation
ii. input
iii. weight/parameter
iv. bias

Options:

a. o= input, f=activation, W= weight/parameter, x= bias
b. f=input, W=activation, b=bias, x= weight/parameter
c. f= activation, W= weight/parameter, x= input, b= bias
d. o=input, f=activation, W= weight, b= bias

III. Fill up the blanks: [Marks = 0.5 X 8 = 4]

1. In the context of neural networks, __________refers to the process of generating new data points from an existing dataset using the learned patterns.
2. The _________is responsible for generating fake data samples, while the _________aims to discriminate between real and fake samples in Generative Adversarial Networks (GANs).
3. In an LSTM cell, the __________determines how much of the previous internal state should be retained.
4. The _________is a measure of how well a neural network's predictions match the desired outputs for a given set of inputs.
5. In the context of neural networks, _________is the process of updating the model's parameters using the gradients of the loss function with respect to the parameters.
6. Recurrent Neural Networks (RNNs) are a type of neural network architecture designed to process _________data by maintaining an internal state or memory.
7. _________is a regularization technique used to prevent the weights of a neural network from becoming too large by adding a penalty term to the loss function.
8. The ___________activation function is commonly used in CNNs to introduce non-linearity into the network and help with the convergence of the training process.

IV. Answer the subsequent questions: [Marks = 1 X 5 = 5]

1. Explain when you would use L1 regularization instead of L2 regularization.

Ans:

2. Neural networks are universal approximators. Give one reason why a neural network may nonetheless fail to approximate a function.

Ans:

3. Describe one thing that makes learning different from pure optimization

Ans:

4. Explain why we do not typically worry about a sigmoid unit saturating when it is used as an output layer.

Ans:

5. If you are using dropout while training a neural network with n layers that are each m units wide, and you take s gradient steps during training, how many random numbers will you have to generate for the dropout?

Ans:

HAPPY LEARNING!