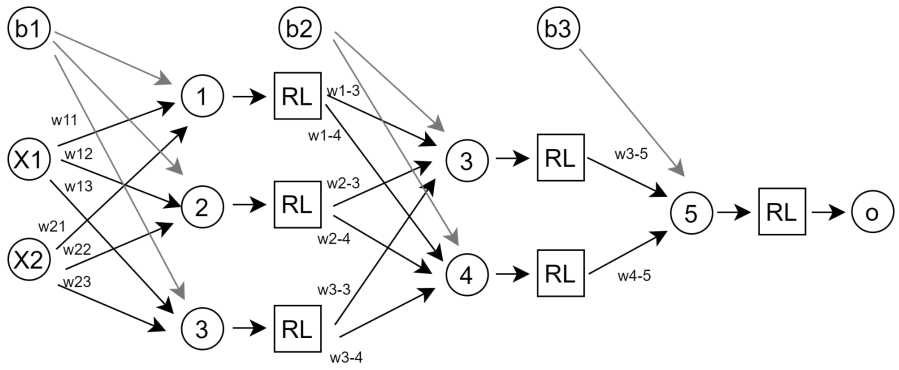
Study Exercises: Concepts of Neural Networks

1. What is the fundamental difference between linear learning and non-linear learning?
2. What are the advantages of linear learning approaches in machine learning and data analysis?
3. What are the limitations of linear learning methods when compared to non-linear learning methods?
4. What is the relationship of neural networks to the human brain ?
5. What is a neuron ?
6. Can you explain the concept of feature engineering in the context of linear and non-linear learning? How does it differ for each?
7. What is regression from images, and how does it differ from image classification?
8. Can you provide examples of real-world applications for regression based on images ?
9. Can you provide examples of real-world applications for image classification ?
10. How can you represent class labels in image classification tasks?
11. What are some common evaluation metrics for classification models?
12. How do you choose the appropriate loss function for classification tasks?
13. What advantages does a multilayer neural network offer over a single-layer network for complex tasks?
14. Is there a difference between Cost function and Loss function ?
15. What is the primary purpose of a cost function in a neural network?
16. Can you provide examples of scenarios where mean squared error (MSE) is a suitable cost function, and when it is not recommended?
17. How does cross-entropy loss function work for classification problems, and why is it preferred over other loss functions like mean squared error?
18. How does backpropagation work in training a multilayer neural network, and why is it necessary
19. What is the purpose of hidden layers in a neural network?
20. How do hidden layers contribute to the representational power of a neural network?
21. Can you explain the concept of feature extraction in the context of hidden layers?
22. How does the number of neurons within each hidden layer impact the network's capacity and performance?

23. What is the role of activation functions, and why are they important?
24. What are some common activation functions used in hidden layers, and how do they differ in terms of behavior?
25. How do you choose the appropriate number of hidden layers and neurons for a specific problem, and are there any guidelines for doing so?
26. What is feed forward and how does it work ?
27. Can you write high-level instructions on how a training loop should be implemented ?
28. What is an epoch ? How to define the best number of epochs to use ?
29. If you're using gradient descent with a fixed learning rate of 0.01, and you find that the loss is oscillating during training, how might you adjust the learning rate to improve convergence?
30. If you have a neural network with an input layer of 512 neurons, a hidden layer with 256 neurons, and a hidden layer with 128 neurons, what is the size of the output from the last hidden layer before the final output layer?
31. How to prove that your regression or classification network was the best configuration you were able to achieve ?
32. In a feedforward neural network, you have an input of [0.5, 0.3, 0.2] and corresponding weights of [0.2, 0.4, 0.1] for a neuron in the first hidden layer. Calculate the weighted sum for this neuron.
33. Using the weighted sum from the previous question, apply the ReLU activation function. What is the output of the ReLU activation for this neuron?
34. In a classification task with three classes, you have the outputs from the final layer before softmax as [2.1, 1.5, 0.8]. Calculate the softmax probabilities for each class.
35. If you have a feedforward neural network with an input layer of 256 neurons, two hidden layers with 128 and 64 neurons respectively, and an output layer with 10 neurons, how many total parameters (weights and biases) are there in the network?
36. If you have a dataset of 1,000 images and you're using a batch size of 32, how many iterations will it take to complete one epoch (cycle through the entire dataset)?
37. Suppose you have a simple regression model with the following predictions for a set of data points: True values (y): [10, 20, 30, 40, 50], Predicted values ($\hat{y}$): [12, 18, 32, 38, 48] Calculate the Mean Squared Error (MSE) as the loss function for this regression model, and explain what this value signifies in the context of the model's performance
38. Consider a neural network with an input layer, a hidden layer H1 with 8 neurons followed by a ReLU, a hidden layer H2 with 4 neurons followed by a ReLU, an output layer with 1 neuron followed by a sigmoid. Draw the diagram for this small network. This network is best suited for regression or classification ? Why ?

39. Consider a feedforward neural network with 3 layers H1, H2, H3. These layers have 3,2,1 neurons respectively, each neuron has a ReLU activation function. Mean Squared Error (MSE) is used as the loss function. For a specific input (X1=2, X2=3), the weights and biases are as follows:



- Weights for the connections from input to hidden layer:
  - H1 (w11, w12, w13, w21, w22, w23): [0.5, -0.2, 0.8, -0.6, 0.3, 0.1]
  - H2 (w1-3, w1-4, w2-3, w2-4, w3-3, w3-4): [0.1, 0.3, -0.2, -0.1, 0.1, 0.7]
  - H3 (w3-5, w4-5): [0.5, 0.3]
- Biases for the hidden layers H1, H2, H3 (b1, b2, b3): [0.1, -0.3, 0.2]

Calculate the following:

- The weighted sum and activation output of each neuron in the hidden layers for the given input.
- The final prediction made by the network for the given input.
- The MSE loss for this prediction when compared to the true target value, which is 2
- This network is best suited for regression or classification ? Why ?

40. Let's continue with the neural network described in the previous question, which has three hidden layers (H1, H2, H3) with ReLU activation and uses Mean Squared Error (MSE) as the loss function. We'll focus on the backpropagation process for updating weights and biases after computing the loss for a specific input (X1=2, X2=3). Given the following information:

    Predicted output of the network ($\hat{y}$): 3.5
    - True target value (y): 5.0
    - Learning rate (α): 0.01

    Perform the following calculations:
    - Calculate the error (δ) at the output layer (H3).
    - Calculate the gradient of the loss with respect to the weights and biases at the output layer (dw3-5 and db3).
    - Propagate the error δ back to the hidden layers (H2 and H1) and calculate the gradients of the loss with respect to the weights and biases at the hidden layers (dw1-3, dw1-4, dw2-3, dw2-4, dw3-3, dw3-4, db1, db2, db3).

    Note: You can use the ReLU derivative as needed (ReLU'(x) = 1 if x > 0, 0 otherwise).