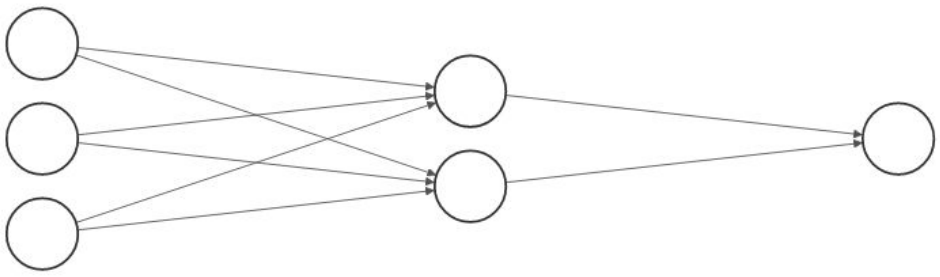
CMPN450 Pattern Recognition and Neural Networks
Sheet 5
Neural Networks

1) Consider the neural network architecture shown in the figure below. The given neural network consists of an input layer, a hidden layer and an output layer. If the weights are randomly initialized such that W[1] = [0.1, 0.2, 0.3; -0.1, 0, -0.2] and W[2] = [0.2 -0.2]. Assume the bias units are zero in both layers. Assume the activation non-linear function used in the hidden layer and the output layer is the sigmoid function.

   (a) Derive the equations of forward propagation of the network.
   (b) Derive the loss function (cost function) in the case of binary classification.
   (c) Derive the equations of backward propagation of the network.

   Suppose the input X = [3; 0.2; -0.1] was fed to the neural network shown above:

   (d) What will be the **output of the network**? Perform one pass of forward propagation to figure out the answer.
   (e) What will be the **prediction (classification) of the network** for the same given input if the threshold applied after the sigmoid activation function in the output layer is **0.6**? Assume Class 1 is chosen if the output of the network is above or equal to the threshold, and class 0 otherwise.
   (f) What will be the actual cost incurred by the network if the actual class is 0 for the given input?
   (g) Perform one pass of backward propagation to find the partial derivatives of the loss function with respect to W1, W2, b1 and b2.
   (h) Perform one pass of gradient descent update to find the new updated set of weights and biases. Use learning rate = 0.05



(2) What would have changed in the previous question if the neural network was solving a **regression problem** rather than **a classification problem**? Write down the modified equations.

(3) What would have changed in the previous question if the neural network was solving a **multi-class classification problem** rather than **a binary classification** problem? Write down the modified equations.

(4) What does the **learning rate** (alpha parameter) mean? How does changing the learning rate affect the training process?

(5) What is the difference between **batch gradient descent** and **sequential (stochastic) gradient descent**?

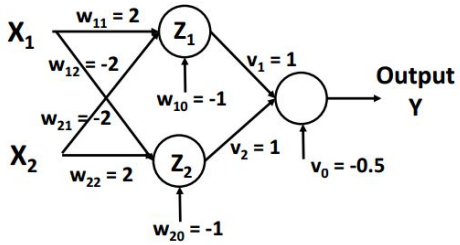
(6) Mention four different types of **activation functions**. Write down the mathematical expression for each of them as well as their **derivatives**. Mention the advantage(s) and disadvantage(s) of each of them.

(7) What are the **hyperparameters** in the gradient descent update algorithm? How to select these hyperparameters?

(8) What is the difference between **training set**, **test set** and **validation set**? Are there any guidelines in selecting each of them?

(9) Mention the difference between **overfitting** and **underfitting**? Give an example to each of them.

(10) What is the output of the following neural network? Given that the non linear activation function is a unit step function? Assume X1 and X2 are binary inputs.



(11) Modify the weights of the previous network so it can compute the XNOR function?

(12) What are different optimization algorithms? State the weight update equation for each of these optimizers.

(13) What is the main difference between gradient descent and gradient descent with momentum?

(14) Mention two different ways to reduce overfitting. Explain how each of them reduces overfitting.

(15) Choose the correct answer:

1. Which hyperparameter of the following needs to be tuned first in a typical neural network problem?
    a. Momentum parameter.
    b. Mini batch size
    c. Learning rate
    d. Number of hidden nodes in each layer.

2. The softmax layer is only used in
    a. Binary classification problems.
    b. Multiclass classification problems.
    c. Regression problems.
    d. All of the above.

3. As the number of hidden layers increases in a neural network,
    a. The time needed to train the network decreases.
    b. The network learns more complex functions and features.
    c. The network converges to a local minimum of the cost function faster.
    d. The network may be subject to overfitting.

4. The activation function recommended to use when working with images is
    a. ReLU function.
    b. Step function.
    c. Sigmoid function.
    d. Tanh function.

5. The neural network that tries to match two given inputs and detect how similar or different they are from each other is called
    a. Convolutional neural network.
    b. Siamese network.
    c. Recurrent neural network.
    d. Generative Adversarial Network.

6. A network with a skip connection from output layer to input layer is called
    a. Convolutional neural network.
    b. Siamese network
    c. Recurrent Neural Network.
    d. Generative Adversarial Network.

7. A neural network used mainly to generate features from input images and represent these images in a compressed low dimensional space is called:
    a. Convolutional neural network.
    b. Auto Encoder network.
    c. Recurrent Neural Network.
    d. Siamese Network.

8. If the neural network is subject to overfitting, then we can reduce the effect of overfitting by:
    a. Increasing the size of the training data.
    b. Increasing the size of the neural network.
    c. L2- Regularization.
    d. Dropout regularization.

9. A time series is composed of:
    a. Trend
    b. Seasonality
    c. Random Noise
    d. All of the above

10. For the neural network to learn functions such as XOR and XNOR, it is sufficient to have:
    a. 1 input layer and 1 output layer.
    b. 1 input layer, 1 hidden layer, 1 output layer.
    c. 1 input layer, 2 hidden layers, 1 output layer.
    d. It is dependent on the number of inputs, and so it is impossible to tell.