QUESTION 1 (31/3)

Which of the following statements are true? You may choose multiple answers.

- [ ] A convolution neural network (CNN) is more suitable for image processing applications than sequential data processing applications.
- [ ] A recurrent neural network (RNN) is more suitable for image processing applications than sequential data processing applications.
- [ ] A convolution neural network (CNN) is more suitable for sequential data processing applications than image processing applications.
- [ ] A recurrent neural network (RNN) is more suitable for sequential data processing applications than image processing applications.

QUESTION 2 (31/3)

Which of the following statements are true? You may choose multiple answers.

- [ ] If a neural feedforward network has 2 hidden layers, it is likely to have a serious gradient vanishing problem.
- [ ] The max. value of the gradient of a sigmoid activation function is 0.25, thus the vanishing gradient problem will occur if a neural net involves many stages of the sigmoid activation function.
- [ ] The tanh activation function can produce negative output responses.
- [ ] The gradient of a ReLU (Rectified Linear Unit) activation function is 1.

QUESTION 3 (31/3)

A recurrent neural network (RNN) is shown in the diagram below.

The weights used in the recurrent neural network (RNN) are whx, whh.

%assume whx,whh,why are initialized at t=1 as

whx=[0.30 0.12 0.24 0.14

0.62 0.26 0.45 0.32]

whh =[0.21 0.32

0.22 0.38]

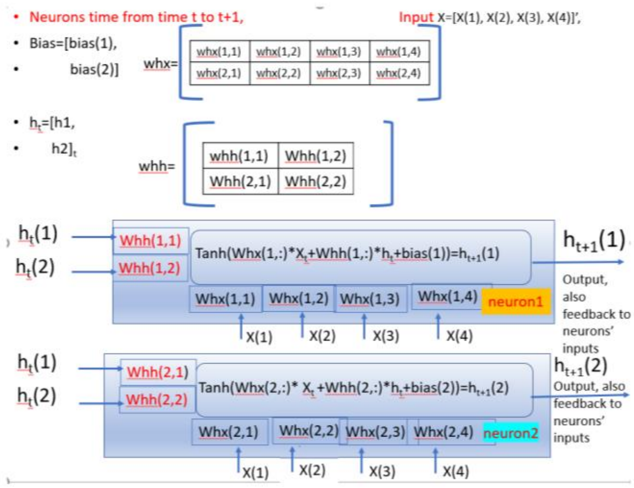
bias=[0.13, 0.34]';% initialized bias

ht=1 =[0.11, 0.23]'; %Initialized h

Note:

The input at time t=1 is [X(1), X(2), X(3), X(4)]= [1,0, 0, 0]

Find the output h $_{t=2}$(2).



QUESTION 4 (31/3)

The output circuit of a recurrent neural network (RNN) is as follows.

If ht(1)=0.15, ht(2)=0.36, ht(3)=0.14

why=[0.24 0.27 0.45

0.05 0.08 0.21

0.35 0.26 0.64

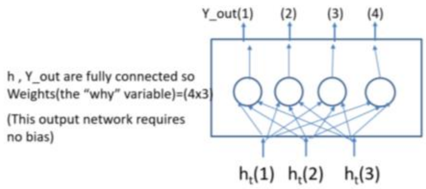
0.76 0.13 0.71]

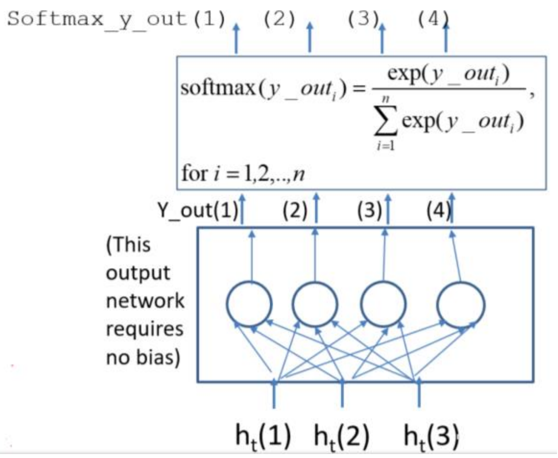
Find Softmax_y_out(2).

why=

why(1,1)	why(1,2)	why(1,3)
why(2,1)	why(2,2)	why(2,3)
why(3,1)	why(3,2)	why(3,3)
why(4,1)	why(4,2)	why(4,3)

- Y_out(1)=why(1,1)*ht(1)+why(1,2)*ht(2)+ why(1,3)*ht(3)
- Y_out(2)=why(2,1)*ht(1)+why(2,2)*ht(2)+ why(2,3)*ht(3)
- Y_out(3)=why(3,1)*ht(1)+why(3,2)*ht(2)+ why(3,3)*ht(3)
- Y_out(4)=why(4,1)*ht(1)+why(4,2)*ht(2)+ why(4,3)*ht(3)





QUESTION 5 (27/3)

In the following diagram, it shows the parameters of a part of a neural network at time k. The activation function of the neurons is sigmoid. The energy to be minimized during training is E = (1/2)*(y-t)^2. such that new_w = old_w + Dw, where

Dw = -learning_rate*(∂E/∂w), and Dw=delta_weight

Assume all the weights will be updated together only after all delta weights (Dw) have been calculated for each epoch time k. So use the current (time k) parameters for this calculation.
The current parameters at time k are:

learning_rate=0.5,

x1=0.5, x2=0.8, x3=0.7,

wh1=0.24, wh2=0.43, wh3=0.52,

ow11=0.12, ow12=0.42, ow13=0.1,

bh1=0.33, bh2=0.21, bh3 = 0.15,

s2=0.2, s3=0.1 %are sensitivities

If the new wh1 at time k+1 is 0.234 , find s1.

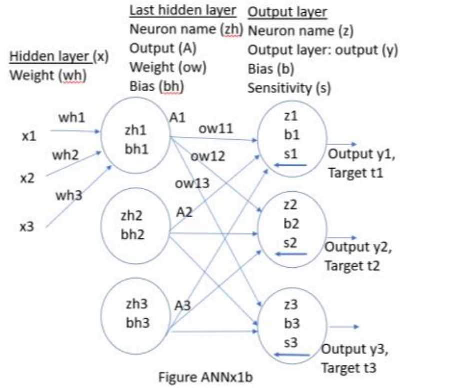


Figure ANNx1b

QUESTION 6 (27/3)

The input layer to the first convolution layer feature maps (C1) of a convolution neural network CNN is shown.

Input is 128 x 128

kernel size=5x5

Step size=1

For this part of the neural network, the number of weights is N and the number of biases required is B

Find N+B.

QUESTION 7 (27/3)

In a Convolution Neural Network CNN , layer C(i) has 64 feature maps, each feature map is of size 7x7. This feature map (layer C(i)) is fully connected to the next layer F(i+1) with 80 neurons. Find the number of weights between C(i) and F(i+1).

QUESTION 8 (27/3)

As shown in the diagram, a 3x3 window (WinA) is selected from the input image.

Using the CNN model, the parameters are
WinA=[0.2 0.23 0.31
0.21 0.14 0.32

0.33 0.14 0.54]

K'=[0.3 0.2 0.25
0.33 0.13 0.22
0.53 0.11 0.4]

The convolution result of WinA and the 3x3 Kernel (K) is connected through a sigmoid activation function to a feature in the feature map.

If the bias of the kernel is 0.5, find the output value of the feature output (FeatA) of the window (WinA) at the feature map. Assume K' is the flipped (up-side-down and left-right) version of Kernel K, so, just use correlation of WinA and K' to find the feature map output.