Deep Learning Foundation and Application (AI61002)
End Semester Examination, Spring 2022

Instructions

- Exam Time: 8:00 AM -10:00 AM
- Total Questions: 52
- Marks per question: 48x2, 4x1
- Total Marks: 100
- **ALL QUESTIONS ARE MANDATORY.**
- **No negative marks.**
- **The exam portal will be closed at 10:00 AM.**
- **For numerical answers DO NOT add any ',' or other special characters in the answer.**
- **When the format for answering a question is specified, STRICTLY follow the format to enter the answer.**

1

Which CNN Architecture below uses 1x1 convolutions to reduce the number of parameters of the model? *
(2 Points)

- AlexNet
- VGG

- [x] GoogleNet
- [ ] All of the above

2

In a 2-D convolutional neural network having an input with 2 channels, what should be
the dimensionality of each convolutional kernel? *
(2 Points)

2

3

State whether true or false. (Write 1 for true and 0 for false.)

Local response normalization is a non-trainable layer and batch normalization is a trainable layer. *
(2 Points)

1

4

How many learnable parameters does a zero bias neural network with 20 input neurons; 10 hidden neurons and 2 output neurons have? *
(2 Points)

220

5

For a network A with a depth of 200, if the number of operations are 12 Billion. Compute the approximate number of operations in the order of Billions approximated to 2 decimal places for a similar network B with a depth of 152. *
(2 Points)

9.12

6

RNN remembers each and every information through ____.

*
(2 Points)

- Time
- Work
- Memory
- None of these

7

An image of size 3x196×196 is transformed to 3x21×21 by a max-pooling layer of size 6×6 with stride 5 and padding 10. Calculate the number of weights that are needed to be learned? *
(2 Points)

0

8

Calculate the number of trainable parameters for the convolutional neural network given below: (Input size is represented as H x W x C and consider bias for all layers.) *

(2 Points)

(input) 28×28×3 → (conv) 12c3w1s0p → (maxpool) 2w2s0p → (conv) 6c3w2s
→ (*maxpool*) 2w2s0p → (flat) 54 → (fc) 20 → (output) 10

2300

9

If the depth of a network A is increased by a factor α and its width is increased by a factor of β, resulting in a network B. However, it is found that the number of operations, given by OPS(.), of A and B are same, i.e., OPS(A)=OPS(B). Find the relation between α
and β:
*

(2 Points)

- α=√β
- α=(1/√β)
- β=(1/√α)
- β=√α

10

What is the use of 'model.zero_grad()' for a network named 'model' in PyTorch? *

(2 Points)

- It initializes all parameters of the network to zero.

- (●) It initializes the gradient buffers of all parameters of the network to zero
- ( ) It computes the gradient of the first layer with respect to the weights.
- ( ) None of these.

11

How many unique equations are needed to represent the ML system given by the equation below ? *

(2 Points)

$$y = f(x) = ax_1 + bx_2 + cx_3 +$$

4

12

If a 3x28×28 tensor is flattened to one dimension, then what is the length of the resultant tensor? *

(2 Points)

2352

13

Calculate the operational compute complexity of the fully-connected neural network (without bias) given below when operating it in forward pass:

(input) 200 -> (fc) 100 -> (fc) 25 -> (fc) 10

*
(2 Points)

22750

14

Consider a multi-layer perceptron network with 200 hidden layers where sigmoid activation function is applied in each layer. Which of the following layers is/are most affected by vanishing gradients? *
(2 Points)

- Output layer
- Intermediate layers
- Non-differentiable loss block
- Initial Layers

15

1x1 convolutions are used for _____. Select the correct options from below to complete the statement. *
(2 Points)

- learning local spatial features

learning global spatial features

channel mixing

All of the above

16

Read the question below and choose an option. you have to write only one option. example: if your answer is option b then write **b** only, don't write anything else with the option. *
(2 Points)

Which of the following equations define Margin Loss cost function? (Consider o_k=outpu the neural network, t_k= target response desired, M=Margin criteria and K= Total numbe classes)

a. $J(.) = -\frac{1}{K}\sum_{k=1}^{K} log(1 + e^{-o_k t_k})$

b. $J(.) = -\frac{1}{K}\sum_{k=1}^{K} w_k y_k$

c. $J(.) = \frac{1}{K}\sum_{k=1}^{K} max(0, M - o_k t_k)$

d. $J(.) = -\frac{1}{K}\sum_{k=1}^{k} |o_k - t_k|$

c

17

A Multi-layer perceptron consists of 50 input neurons, followed by 100 hidden neurons, followed by 10 output neurons with a bias at each layer. If the network weights are represented in double-precision floating-point numbers, how much memory will be required to save the network weights after training? *
(2 Points)

48880 bytes

18

Calculate the number of trainable parameters for the fully-connected neural network (with bias) given as,
(input) 2408 -> (fc) 30 -> (fc) 50 -> (output) 10 *
(2 Points)

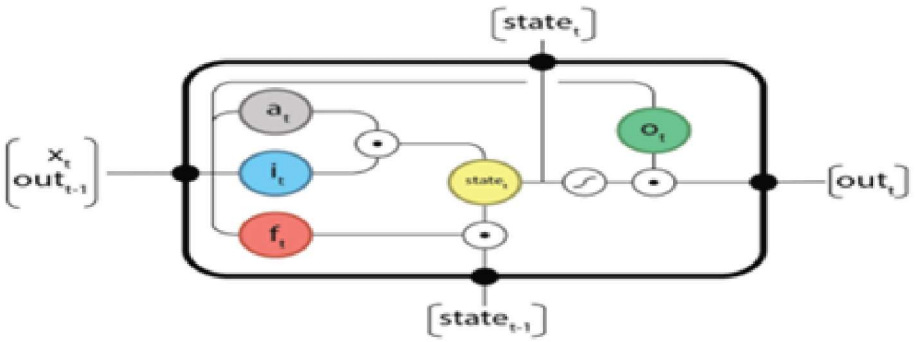
74330

19

Find forget gate (ft)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- ⊙ is the element-wise product or Hadamard product.
- • is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

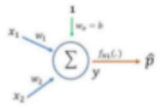
Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

1.7

20

Read the question below and choose an option. you have to write only one option. example: if your answer is option b then write **b** only, don't write anything else with the option. *
(2 Points)

Consider a simple perceptron network having two input neuron and a non-linear transfer function as shown below



The derivative of the cost function of the above network with respect to the weights may be represented as:

$$\frac{\partial J(W)}{\partial W} = \frac{\partial J(W)}{\partial \hat{p}} \frac{\partial \hat{p}}{\partial y} \frac{\partial y}{\partial w}$$

Which of the following is True with respect to the derivative of loss function defined above?

a. $\frac{\partial J(W)}{\partial \hat{p}}$ = derivative of the cost function, $\frac{\partial \hat{p}}{\partial y}$ = derivative of the linear network, $\frac{\partial y}{\partial w}$ = derivative of non-linear transfer function

b. $\frac{\partial J(W)}{\partial \hat{p}}$ = derivative of the linear network, $\frac{\partial \hat{p}}{\partial y}$ = derivative of the cost function, $\frac{\partial y}{\partial w}$ = derivative of non-linear transfer function

c. $\frac{\partial J(W)}{\partial \hat{p}}$ = derivative of the cost function, $\frac{\partial \hat{p}}{\partial y}$ = derivative of non-linear transfer function, $\frac{\partial y}{\partial w}$ = derivative of the linear network

d. $\frac{\partial J(W)}{\partial \hat{p}}$ = derivative of the linear network, $\frac{\partial \hat{p}}{\partial y}$ = derivative of non-linear transfer function, $\frac{\partial y}{\partial w}$ = *derivative of the cost function*

c

21

VGG16 has _____ convolutional layers

*

(1 Point)

13

22

Read the question below and choose an option. you have to write only one option. example: if your answer is option b then write **b** only, don't write anything else with the option. *

(2 Points)

What is the expression for calculating the hidden state in an RNN? Given, h_t denotes a hidden stat time t, h_{t-1} is a previous hidden state, x_t is input at time t, f_{NL} is a non-linear function, W{h,x,t} and W{h,h,t} are the weights represent the input to hidden transformation state in t^{th} dimension and hidden to hidden transformation state in t^{th} dimension respectively.

a. $h_t = f_{NL}(W_{h,x,t} . x_t + W_{h,h,t} . h_{t-1})$

b. $h_t = f_{NL}(W_{h,x,t} . h_{t-1} + W_{h,h,t} . x_t)$

c. $h_t = f_{NL}(W_{h,x,t} . h_{t-1} + W_{h,x,t} . x_t)$

d. $h_t = f_{NL}(W_{h,h,t} . h_{t-1} + W_{h,h,t} . x_t)$

a

23

Compute the dimensions (rows×columns) of encoder weight matrix of an autoencoder having 512 hidden units and input size of 28. Assume input and activations to be column vectors.
Part B: What is the value of columns? *
(1 Point)

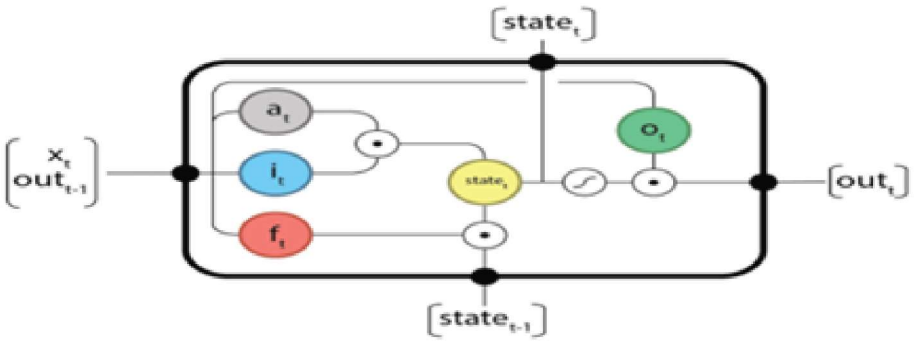
29

24

Find output gate (ot)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- ⊙ is the element-wise product or Hadamard product.
- • is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

1.65

25

You have a CNN with an input Image of size 3x8x8 and a convolution kernel of size 3x5x5. The kernel has a bias associated with it. What will be the total number of operations (without FMA) if you perform convolution with a single 3x5x5 kernel? *

(2 Points)

2400

26

Which algorithm is used to train recurrent neural networks?

*

(2 Points)

- Back-propagation
- Forward-propagation through time
- Back-propagation through time
- All of these.

27

You have an image of size 3x8x8 and you are asked to perform convolution with 5 kernels of size 3x3x3. What will be the number of channels in the intermediate representation obtained? *

(2 Points)

5

28

Which of the following is TRUE about Decoder in adversarial autoencoder? *

(2 Points)

- It generates an image which is different from the initial representation.

It differentiates between real image and generated image.

It generates an image from a given latent space representation.

All of these.

29

Consider a convolutional layer with 10 kernels with height and width 2. If the layer operates on a grayscale image of size 1×132×132 with no padding=5 and stride=7, if the size of the output is represented as C×M×N, what will be the values of C,M and N?

(Write your answer only in the form of C=x, M=y, N=z) *

(2 Points)

C=10, M=21, N=21

30

In a 3-D convolutional neural network having an input volume with only one channel,
what should be the dimensionality of each convolutional kernel? *

(2 Points)

3

31

What will be the width and height of the output after the following operations?

Input size = [3 x 227 x 227], kernel size = [3 x 11 x 11], Stride = 4 and no padding.

Assume no. of kernels to be 5.

(Write your answer only in the form of width=x, height=y) *
(2 Points)

width=55, height= 55

32

Identify the correct usage of SGD optimizer in PyTorch for training a network named 'model'? *
(2 Points)

- optimizer = optim.SGD(model.parameters(), lr=1e-4, momentum = 0.9)
- optimizer = optim.SGD(model, lr=1e-4)
- optimizer = optim.SGD(model, lr=1e-4, momentum = 0.9)
- optimizer = optim.SGD(model.parameters(), lr=0)

33

In 1-D discrete convolution, two input sequences have lengths of 5 and 3 respectively.
What is the output sequence length? *
(2 Points)

7

34

The input to a convolutional kernel has 3 channels. The output after the 2-D convolution operation has 5 channels. How many convolutional kernels are there? *
(2 Points)

5

35

It is given that:

(Input) 3 x 227 x 227 → (Conv2D) 64c 11w 4s 2p → (Output) 64 x56 x 56;

where ‘c’ ‘w’, ‘s’ and ‘p’ represent the channels, filter size, stride, and padding respectively.

Calculate the number of parameters? *
(2 Points)

23296

36

VGG16 has ____ fully connected layers
*
(1 Point)

3

37

For a 2-D convolution, if the output width is the same as input width (stride of 1 and
NO padding is used), what should be the width of the convolutional kernel? *
(2 Points)

1

38

What will happen to the weights of a neural network during training if the learning rate is set to zero? *
(2 Points)

- Weight update will be very slow
- Weight update will tend to zero but not exactly zero
- Weights will be zero
- Weights will not be updated

39

Consider a convolutional kernel of size 1010 applied on an input tensor x with a stride of 1 and padding 1 to get the output tensor y. Which of the following statements is true? *
(2 Points)

- The output tensor y will have lesser number of channels than the input tensor x.
- The dimensions of the output tensor y will be less than the input tensor x.
- The dimensions of the output tensor y will be same as the input tensor x.

Both the dimensions and the number of channels of the output tensor y will change.

40

A feature map containing 8 channels is fed as input to a 2-D maxpool operation. What
is the number of channels in the corresponding output? *
(2 Points)

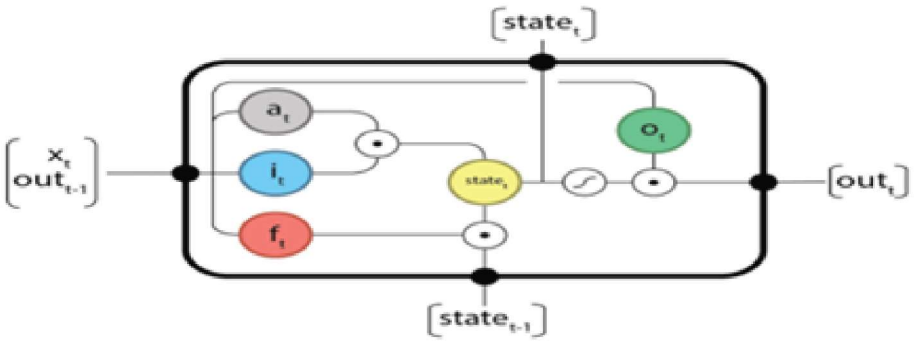
8

41

Find input activation (at)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- ⊙ is the element-wise product or Hadamard product.
- • is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

1.1

42

The input to a convolutional kernel has 4 channels. The output after the 2-D convolution operation has 2 channels. What is the depth (i.e., number of channels) of each
convolutional kernel? *
(2 Points)

4

43

What is the objective for training a neural network? *
(2 Points)

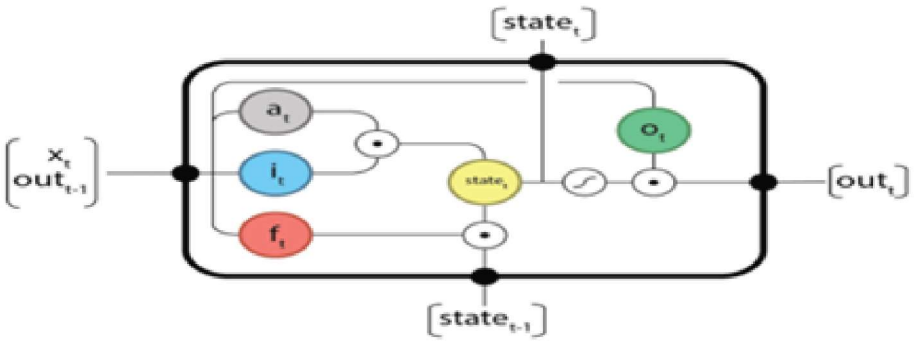
- To change the weights
- To change the biases
- To change both the weights and biases
- To minimize the loss function towards zero

44

Find internal state (st)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- $\odot$ is the element-wise product or Hadamard product.
- $\cdot$ is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

0

45

It is given that:

(Input) 1 x 32 x 32 → (Conv2D) 6c 3w 1s 0p → (Output) 6 x 30 x 30;
where 'c' 'w', 's' and 'p' represent the channels, filter size, stride, and padding respectively.

If the speed is 10x 10^9 flops/sec, what is the throughput? (Use FMA to find the number of operations and round off your answer to the nearest integer.) *
(2 Points)

185185

46

Find the dimension of output(J×K×L) of the convolutional neural network given below,
(Input size is represented as CHW).

(Write your answer in the form of J=x, K=y, L=z)
*
(2 Points)

(input) 3 ×38×38 → (conv) 10c3w5s0p → (maxpool) 2w2s0p → (output) J×K

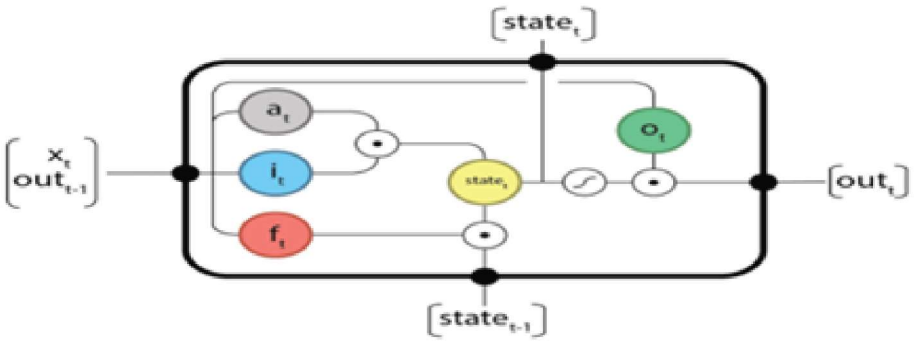
J=10, K=4, L=4

47

Find output (outt)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- ⊙ is the element-wise product or Hadamard product.
- • is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

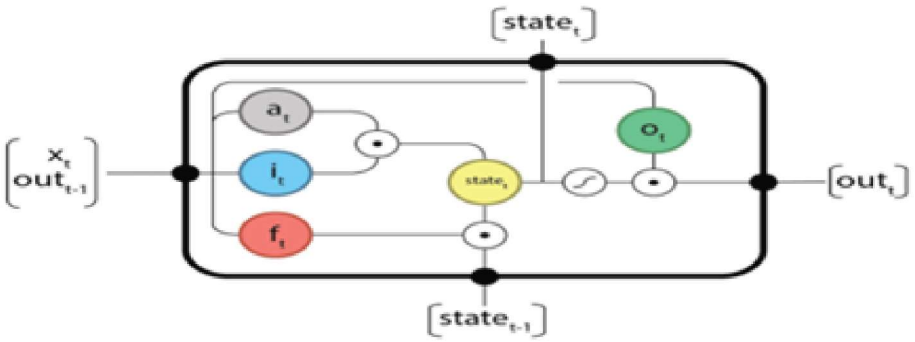
0

48

Find input gate (it)? *
(2 Points)

In the figure given below, the different syntactic notes mentioned are as follows:

- ⊙ is the element-wise product or Hadamard product.
- • is inner product.
- σ represents the sigmoid function.



Internal weights W - weights for input state, U – weights for hidden state are given as

$$W_a = \begin{bmatrix}0.45\\0.25\end{bmatrix}, U_a = [0.15]$$

$$W_i = \begin{bmatrix}0.95\\0.8\end{bmatrix}, U_i = [0.8]$$

$$W_f = \begin{bmatrix}0.7\\0.45\end{bmatrix}, U_f = [0.1]$$

$$W_o = \begin{bmatrix}0.6\\0.4\end{bmatrix}, U_o = [0.25]$$

Input:

$$X_1 = \begin{bmatrix}1\\2\end{bmatrix}$$, assume $out_0 = 0$, $state_0 = 0$

Using this information find the input activation (a_t), input gate (i_t), Forget gate (f_t) and output gate (o_t), Internal state ($state_t$) and output (out_t) at t = 1 in forward pass. Answer should be rounded off t 2 decimal places.

3.35

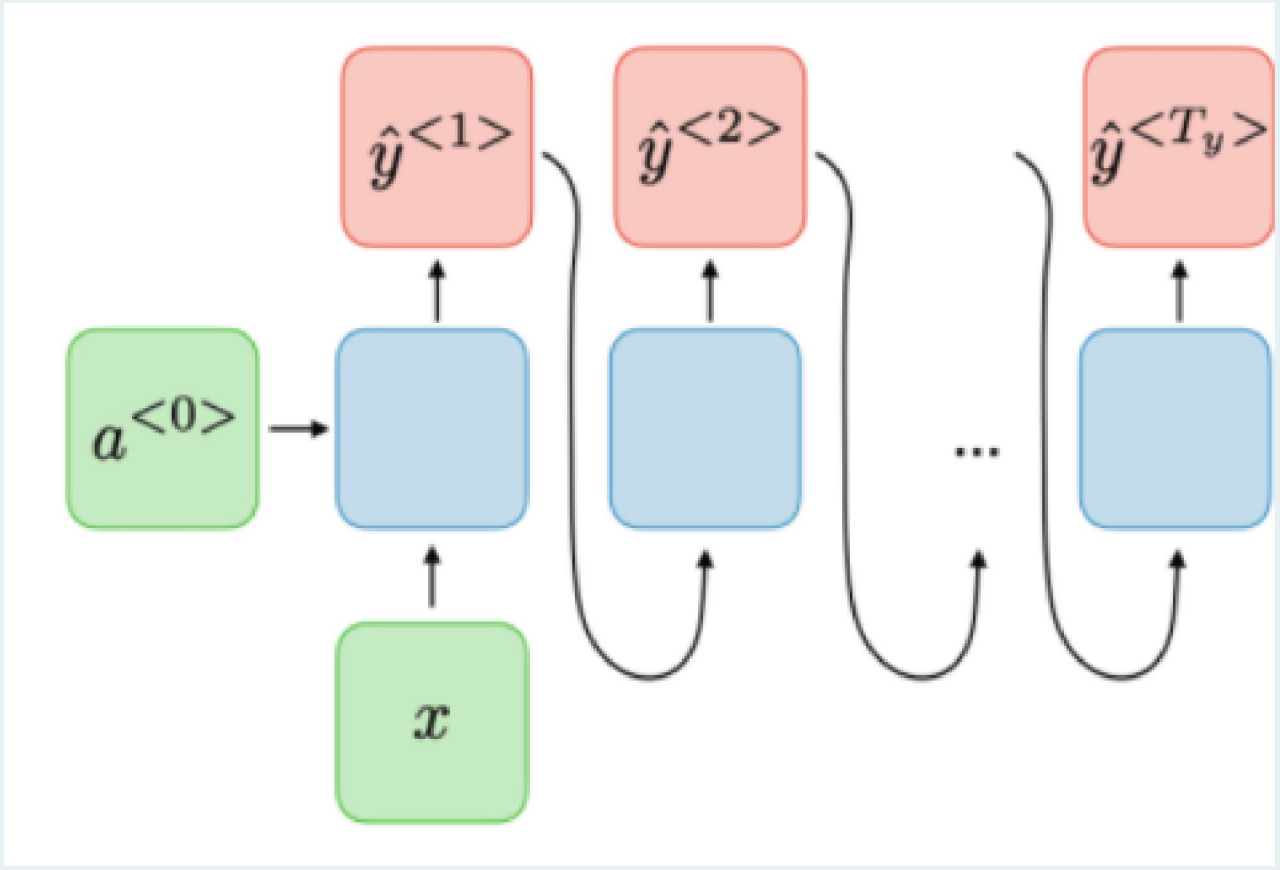
49

Compute the dimensions (rows×columns) of encoder weight matrix of an autoencoder having 512 hidden units and input size of 28. Assume input and activations to be column vectors.
Part A: What is the value of rows? *
(1 Point)

512

50

Which type of RNN is shown in the given figure? *
(2 Points)



- One-to-one
- One-to-many
- Many-to-one
- Many-to-many

51

```
class Network(nn.Module):
    def __init__(self):
        super (Network,self).__init__()
        self.Layer1= nn.Sequential(
            nn.Linear(32*32, 512),
            nn.ReLU(),
            nn.Linear(512,256),
            nn.ReLU())
        self.Layer2 = nn. Sequential(
            nn.Linear(256,20))

    def forward (self, x):
        x= self.Layer1(x)
        x= self.Layer2(x)
        return x
```

Which among the following best defines the network shown below in the code snippet? *
(2 Points)

- Autoencoder Network
- Feed Forward Neural network
- Convolutional Neural network
- None of these

52

For a filter of size 11x11, what should be the padding P on each side such that the output has the same size as the input when the convolution operation is carried with the default stride of the unit? *
(2 Points)

5