Quiz-01

- Due Sep 1 at 11:59pm
- Points 10
- Questions 10
- Available Aug 30 at 12:01am - Sep 1 at 11:59pm
- Time Limit None
- Allowed Attempts 3

Instructions

Intro and Universal Approximators

This quiz covers lectures 1 and 2. Several of the questions invoke concepts from the hidden slides in the slide deck, which were not covered in class. So please go over the slides before answering the questions.

You will have three attempts for the quiz. Questions will be shuffled and you will not be informed of the correct answers until after the deadline. While you may discuss the concepts underlying the questions with others, you must solve all questions on your own - see course policy.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
KEPT	Attempt 1	78 minutes	7.5 out of 10
LATEST	Attempt 2	29 minutes	6.5 out of 10
	Attempt 1	78 minutes	7.5 out of 10

Correct answers are hidden.

Score for this attempt: 6.5 out of 10
Submitted Aug 31 at 4:09pm
This attempt took 29 minutes.

Question 1
1 / 1 pts
Which of your quiz scores will be dropped?

- Lowest 1 quiz scores
- Lowest 2 quiz scores

- Lowest 3 quiz scores
- No scores will be dropped

IncorrectQuestion 2

0 / 1 pts

When was the first connectionist network model proposed?

Slide: lec 1, “Connectionism”

- 1949 by Donald Hebb
- 1873 by Alexander Bain
- 1749 by David Hartley
- 1943 by McCulloch and Pitts
- 1957 by Frank Rosenblatt

See recording of lec 1, “Connectionism” slide for firm confirmation that this was the first connectionist model. The slide text also implies this, albeit a little less clearly.

Question 3

1 / 1 pts

Is the following statement true or false? Hebbian learning allows reduction in weights and learning is bounded.

Slide: lec 1, "Hebbian Learning" : Slides 66 - 69

- True
- False

Slides 67: If neuron x repeatedly triggers neuron y, the synaptic knob (Weight) connecting x to y gets larger. Hence the weight only increases and a mechanism for weight reduction is not given. Also, the upper bound to which the weight increases to is not define in the learning, making it unbounded.

PartialQuestion 4

0.5 / 1 pts

Which of the following statements are true about MLPs? (select all that apply)

Slides: lec 2, "Sufficiency of architecture", "The issue of depth", "How many layers for a Boolean MLP?"

- [x] The VC dimension of an MLP is bounded by the square of the number of weights in the network.

- [x] Even in theory, MLPs may need more than one hidden layer to compose complex decision boundaries of sufficient complexity.
- [x] Deeper networks generally require far fewer neurons than shallower networks to express the same function
- [ ] A network comprising exactly one layer (the output layer) is a Universal Boolean Machine

Question 5

1 / 1 pts

Which statements did David Hartley assert or imply in Observations on Man? (select all that apply)

Slide: lec 1, "Dawn of Connectionism"

- [x] Memories can be stored and can be linked to sensory input.
- [ ] The brain is composed of neurons connected in a network.
- [x] The brain records information as vibrations.
- [x] The brain can receive and process sensory input.

IncorrectQuestion 6

0 / 1 pts

A majority function is a Boolean function of N variables that produces a 1 if at least N/2 of the inputs are 1. Which of the following are true? (select all that apply)

- [ ] A fixed-depth Boolean circuit, comprising only AND, OR and NOT gates, will require $\Omega(\exp(N^\alpha))$ gates to compute the majority function ($\alpha > 0$)
- [x] A single perceptron can compute a majority function.
- [x] The number of gates in the smallest Boolean circuit of AND, OR and NOT gates that computes the majority function is polynomial in N.
- [ ] We will require a multilayer perceptron with $\Omega(\exp(N))$ perceptrons to compute a majority function

A single perceptron can receive all inputs, compute the sum of their values, and use a simple threshold activation of >= N/2. Thus, an MLP is not required.

For Boolean circuits, though, the specific lower bound is $\Omega(\exp(N^{1/(d-1)}))$, where d is the depth of the circuit (Smolensky 1993).

Decent explanation at https://eccc.weizmann.ac.il/report/2019/133/download/

IncorrectQuestion 7

0 / 1 pts

Which of the following are impossible in theory? Assume all networks are finite in size, though they can be as large as needed. (select all that apply)

Hint: (1) The MNIST dataset is finite, (2) The neural network is a universal approximator.

- [x] Using a threshold network with one hidden layer to perfectly classify all digits in the MNIST dataset.
- [ ] Using a threshold network, as deep as you need, to precisely calculate the L1 distance from a point to the origin.
- [x] Using a threshold network, as deep as you need, to determine if an arbitrary 2D input lies within the square with vertices {(1, 0), (-1, 0), (0, 1), (0, -1)}.
- [ ] Using a threshold network with one hidden layer to determine with certainty if an arbitrary 2D input lies within the unit circle.

1. There is a finite number of MNIST digits and a single hidden layer network is a universal approximator, so a finite network can classify the data.

2. A finite network can only approximate a circular decision boundary, and some points will be misclassified.

3. The L1 distance is a continuous function of the input. You cannot model it perfectly with discontinuous functions like the threshold function.

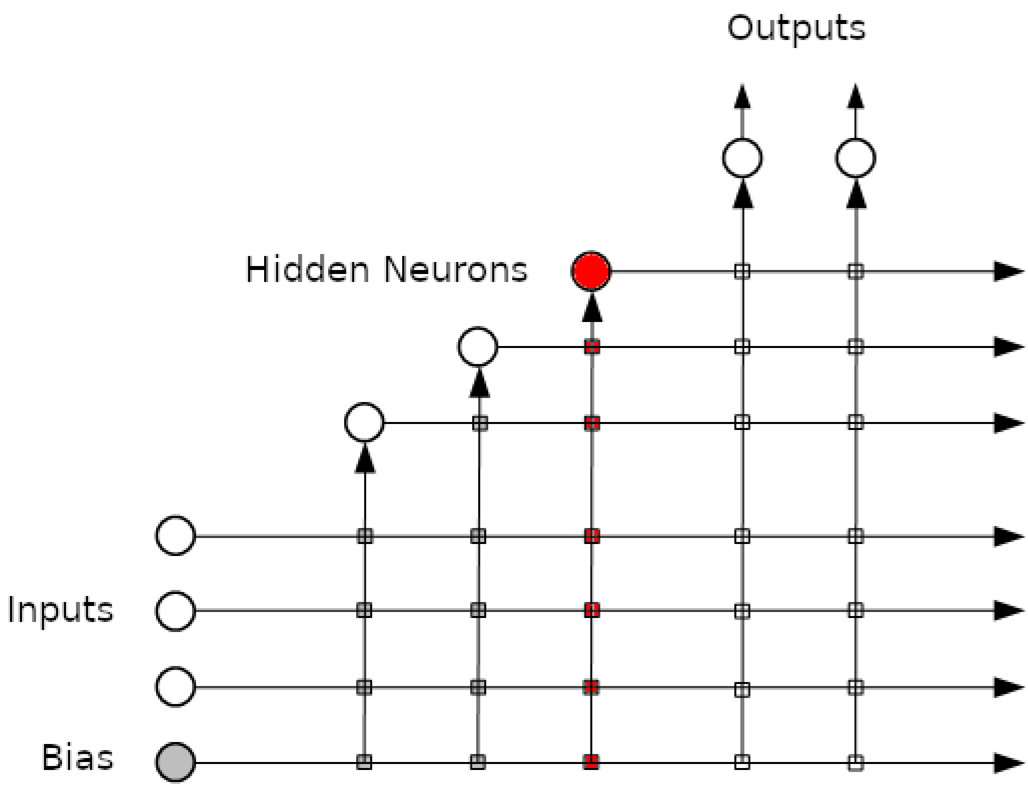
Question 8
1 / 1 pts

The Cascade Correlation architecture (1990, by CMU's Prof. Fahlman!) is relatively unique in that it iteratively modifies its own architecture during training.

It is initialized with no hidden units; to begin, it only has a number of input channels (determined by the dataset) and a number of output units (which may/may not have non-linear activations). This is akin to a single-layer NN.

We then run this training routine:

1. Train output neurons until performance plateaus
2. If error is below some threshold, break
3. Else, freeze ALL network weights. Add a new hidden unit that receives the ORIGINAL input signals AND the outputs of other hidden neurons as inputs.
4. Train this new unit to correlate with the residual errors from previous runs
5. Once adequately trained, attach its outputs to the inputs of the output units. Freeze this unit and unfreeze the output units.
6. Repeat



(img source **https://towardsdatascience.com/cascade-correlation-a-forgotten-learning-architecture-a2354a0bec92** (https://towardsdatascience.com/cascade-correlation-a-forgotten-learning-architecture-a2354a0bec92))

For example, in the diagram above there are 3 original input channels. Each new hidden unit has 3+n input channels, where n is the layer number from 1~N.

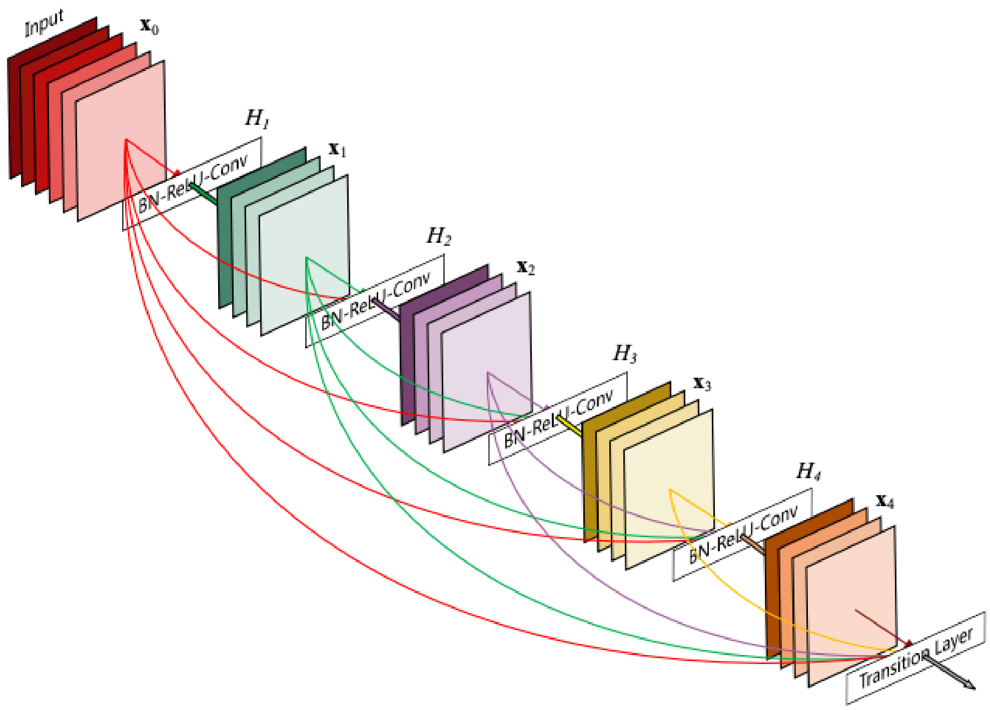
What is the depth of the network above?

(numeric answer, int and float are both fine, also for the definition of network depth, see the lecture 2 recording)

4

Question 9
1 / 1 pts

Densenet (2017) is a CNN network architecture that achieved SoTA results with fewer params than many of its contemporaries. Its main idea was the ‘dense block’: a block where each layer output is concatenated into the input of the downstream layers.



Above is a diagram of a dense-block. Each "H_i" consists of Batchnorm->ReLU->Convolution. Each H_i outputs a "x_i" (one stack of squares in the image) .

What is the depth of the dense block above? (numeric answer, int and float are both fine)

Assume that "BN+RELU+CONV" is a single layer. The square planes simply represent the data flowing down between layers and are not layers themselves.

(Note: for the definition of network depth, see the lecture 2 recording)

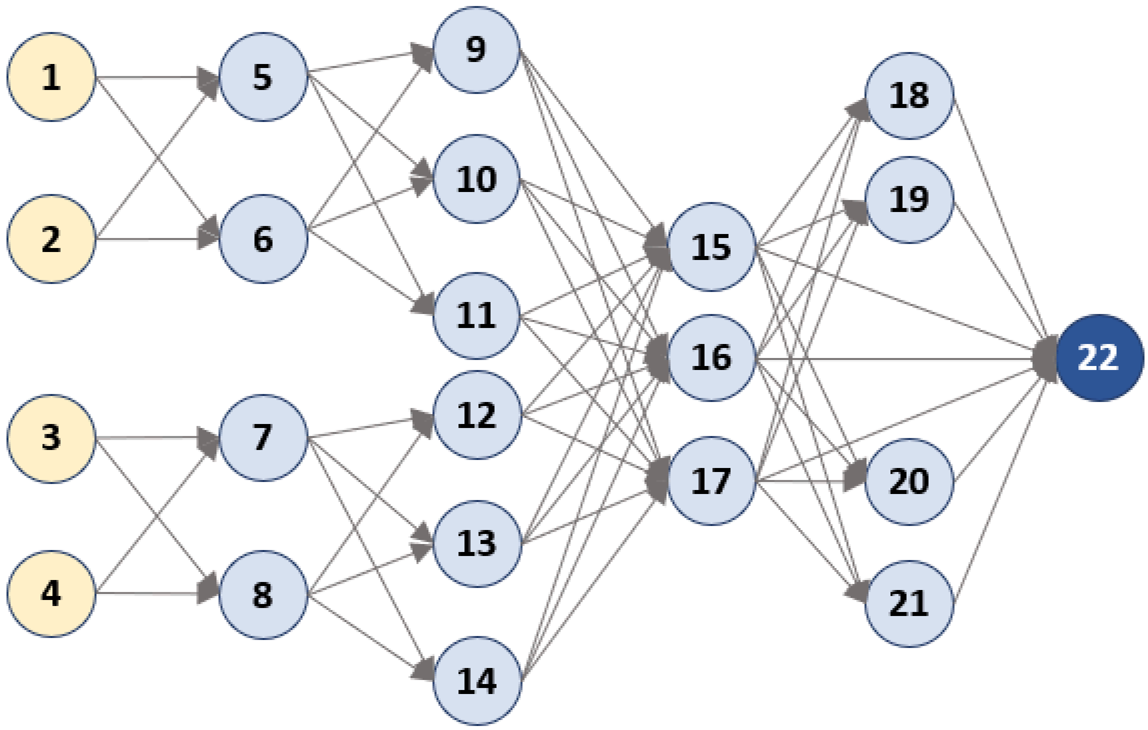
Slide: lec 2, "Deep Structures"

5

The longest path is Input->H_1->H_2->H_3->H_4->Transition: 5 edges.

Question 10

1 / 1 pts



If the yellow nodes are inputs (not neurons) and the dark blue nodes are outputs, which neurons are in layer 4?

(Note: for the definition of network depth and layer number, see the lecture 2 recording)

Slide: lec 2, "What is a layer"

- ○ 15, 16, 17
- ○ 15, 16, 17, 22
- ○ 22
- ◉ 18, 19, 20, 21

22 can't be the 4th layer as it's not visited via 4 edges along the longest path. It's not 15, 16, 17 because they're 3 edges away from the input. 15, 16, 17, 22 wouldn't make sense, as they're not equidistant from the input (22 is one edge farther). That leaves 18, 19, 20, 21. It doesn't matter that they're in a funny configuration; the longest path up to then reaches those four neurons at the same path length. They are layer 4.

Quiz Score: 6.5 out of 10