45 Questions to Test a Data Scientist on Basics of Deep Learning (along with solution)

[Faizan Shaikh](#)
Last Updated : 28 Oct, 2024

15 min read   113

Back in 2009, **deep learning** was only an emerging field. Only a few people recognised it as a fruitful area of research. Today, it is being used for developing applications which were considered difficult or impossible to do till some time back.

Speech recognition, image recognition, finding patterns in a dataset, object classification in photographs, character text generation, self-driving cars and many more are just a few examples. Hence it is important to be familiar with deep learning and its concepts.

In this skilltest, we tested our community on basic concepts of Deep Learning. A total of 1070 people participated in this skill test.

If you missed taking the test, here is your opportunity to look at the questions and check your skill level. If you are just getting started with Deep Learning, here is a course to assist you in your journey to Master Deep Learning:

Table of Contents

1. [Overall Results](#)

2. [Helpful Resources](#)

3. [Questions and Answers](#)

Overall Results

Below is the distribution of scores, this will help you evaluate your performance:
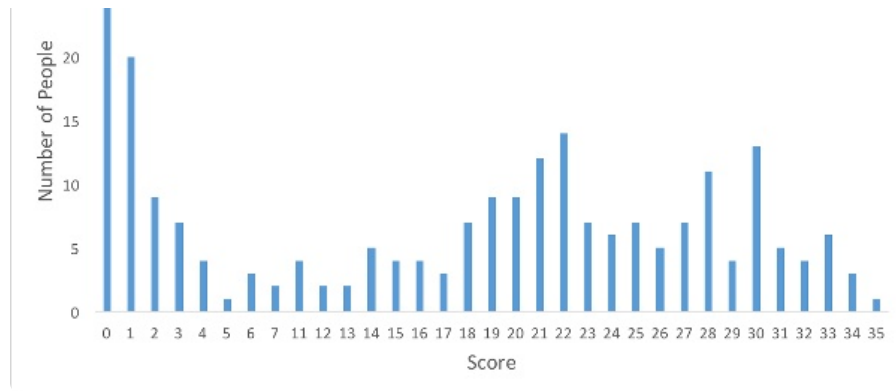
You can access your performance **here**. More than 200 people participated in the skill test and the highest score was 35. Here are a few **statistics** about the distribution.

Overall distribution

Mean Score: 16.45

Median Score: 20

Mode Score: 0

It seems like a lot of people started the competition very late or didn't take it beyond a few questions. I am not completely sure why, but may be because the subject is advanced for a lot of audience.

If you have any insight on why this is so, do let us know.

Helpful Resources

- **Fundamentals of Deep Learning – Starting with Artificial Neural Network**

- **Practical Guide to implementing Neural Networks in Python (using Theano)**

- **A Complete Guide on Getting Started with Deep Learning in Python**

- **Tutorial: Optimizing Neural Networks using Keras (with Image recognition case study)**

- **An Introduction to Implementing Neural Networks using TensorFlow**

Q1. A neural network model is said to be inspired from the human brain.

The neural network consists of many neurons, each neuron takes an input, processes it and gives an output. Here's a diagrammatic representation of a real neuron.
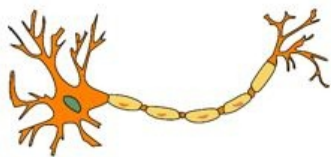

Which of the following statement(s) correctly represents a real neuron?

A. A neuron has a single input and a single output only

B. A neuron has multiple inputs but a single output only

C. A neuron has a single input but multiple outputs

D. A neuron has multiple inputs and multiple outputs

E. All of the above statements are valid

Solution: (E)

A neuron can have a single Input / Output or multiple Inputs / Outputs.

Q2. Below is a mathematical representation of a neuron.
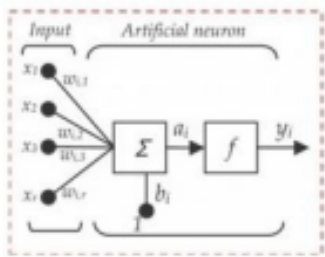
- **x1, x2,..., xN: These are inputs to the neuron. These can either be the actual observations from input layer or an intermediate value from one of the hidden layers.**

- **w1, w2,...,wN: The Weight of each input.**

- **bi: Is termed as Bias units. These are constant values added to the input of the activation function corresponding to each weight. It works similar to an intercept term.**

- **a: Is termed as the activation of the neuron which can be represented as**

- **and y: is the output of the neuron**

$$a = f\left(\sum_{i=0}^{N} w_i x_i\right)$$

Considering the above notations, will a line equation (y = mx + c) fall into the category of a neuron?
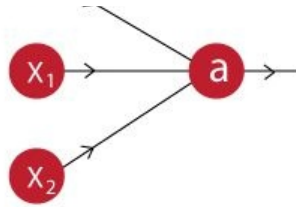
A. Yes

B. No

Solution: (A)

A single neuron with no non-linearity can be considered as a linear regression function.

Q3. Let us assume we implement an AND function to a single neuron. Below is a tabular representation of an AND function:

X1	X2	X1 AND X2
0	0	0
0	1	0
1	0	0
1	1	1

The activation function of our neuron is denoted as:

$$f(x) = \begin{cases} 0, & for\ x < 0 \\ 1, & for\ x \geq 0 \end{cases}$$

What would be the weights and bias?

**(Hint: For which values of w1, w2 and b does our neuron implement an AND function?)**

A. Bias = -1.5, w1 = 1, w2 = 1

B. Bias = 1.5, w1 = 2, w2 = 2

C. Bias = 1, w1 = 1.5, w2 = 1.5

D. None of these

Solution: (A)

A.

1. f(-1.5*1 + 1*0 + 1*0) = f(-1.5) = 0

2. f(-1.5*1 + 1*0 + 1*1) = f(-0.5) = 0

3. f(-1.5*1 + 1*1 + 1*0) = f(-0.5) = 0

4. f(-1.5*1 + 1*1+ 1*1) = f(0.5) = 1

Therefore option A is correct

Q4. A network is created when we multiple neurons stack together. Let us take an example of a neural network simulating an XNOR function.
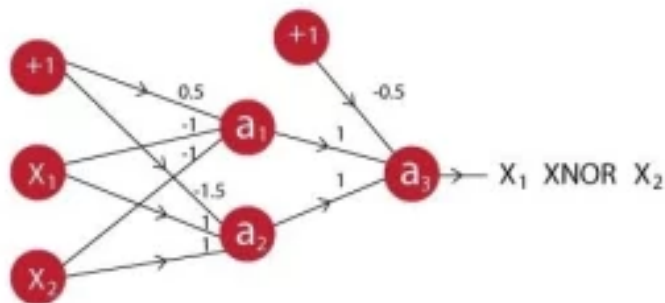
$$f(x) = \begin{cases} 0, & for \ x < 0 \\ 1, & for \ x \geq 0 \end{cases}$$

Suppose X1 is 0 and X2 is 1, what will be the output for the above neural network?

A. 0

B. 1

Solution: (A)

Output of a1: f(0.5*1 + -1*0 + -1*1) = f(-0.5) = 0

Output of a2: f(-1.5*1 + 1*0 + 1*1) = f(-0.5) = 0

Output of a3: f(-0.5*1 + 1*0 + 1*0) = f(-0.5) = 0

Solution (A)

So the correct answer is A

Q5. In a neural network, knowing the weight and bias of each neuron is the most important step. If you can somehow get the correct value of weight and bias for each neuron, you can approximate any function. What would be the best way to approach this?

A. Assign random values and pray to God they are correct

B. Search every possible combination of weights and biases till you get the best value

C. Iteratively check that after assigning a value how far you are from the best values, and slightly change the assigned values values to make them better

D. None of these

Solution: (C)

Option C is the description of gradient descent.

Q6. What are the steps for using a gradient descent algorithm?

1. **Calculate error between the actual value and the predicted value**

2. **Reiterate until you find the best weights of network**

3. **Pass an input through the network and get values from output layer**

~~5. Go to each neurons which contributes to the error and change its respective values to reduce the error~~

A. 1, 2, 3, 4, 5

B. 5, 4, 3, 2, 1

C. 3, 2, 1, 5, 4

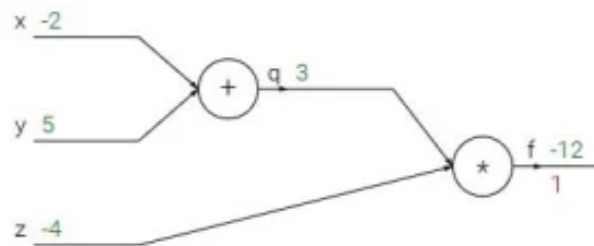D. 4, 3, 1, 5, 2

Solution: (D)

Option D is correct

Q7. Suppose you have inputs as x, y, and z with values -2, 5, and -4 respectively. You have a neuron 'q' and neuron 'f'
with functions:

q = x + y

f = q * z

Graphical representation of the functions is as follows:

What is the gradient of F with respect to x, y, and z?

(HINT: To calculate gradient, you must find (df/dx), (df/dy) and (df/dz)) A. (-3,4,4)

B. (4,4,3)

C. (-4,-4,3)

D. (3,-4,-4)

Solution: (C)

Option C is correct.

- A neural network is a (crude) mathematical representation of a brain, which consists of smaller components called neurons.

- Each neuron has an input, a processing function, and an output.

- These neurons are stacked together to form a network, which can be used to approximate any function.

- To get the best possible neural network, we can use techniques like gradient descent to update our neural network model.

Given above is a description of a neural network. When does a neural network model become a deep learning model?

A. When you add more hidden layers and increase depth of neural networkB. When there is higher dimensionality of data

C. When the problem is an image recognition problem

D. None of these

Solution: (A)

More depth means the network is deeper. There is no strict rule of how many layers are necessary to make a model deep, but still if there are more than 2 hidden layers, the model is said to be deep.

Q9. A neural network can be considered as multiple simple equations stacked together. Suppose we want to replicate the function for the below mentioned decision boundary.

h1   **h2**
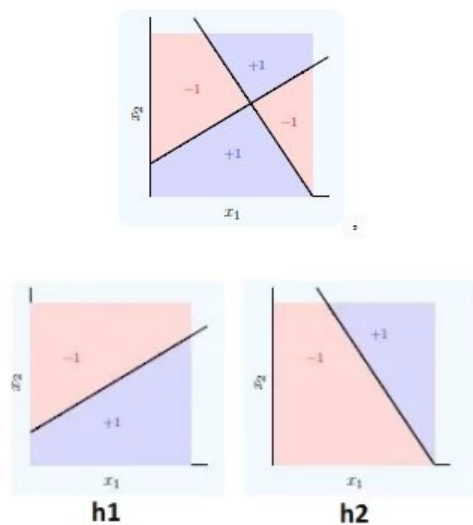
Using two simple inputs h1 and h2

What will be the final equation?

A. (h1 AND NOT h2) OR (NOT h1 AND h2)

C. (h1 AND h2) OR (h1 OR h2)

D. None of these

Solution: (A)

As you can see, combining h1 and h2 in an intelligent way can get you a complex equation easily. Refer Chapter 9 of **this book**

Before checking the next deep learning question, explore the **Certified AI & ML Blackbelt+ Program**!

Q10. "Convolutional Neural Networks can perform various types of transformation (rotations or scaling) in an input". Is the statement correct True or False?

A. True
B. False

Solution: (B)

Data Preprocessing steps (viz rotation, scaling) is necessary before you give the data to neural network because neural network cannot do it itself.

Q11. Which of the following techniques perform similar operations as dropout in a neural network?

A. BaggingB. Boosting

C. Stacking

D. None of these

Solution: (A)

Dropout can be seen as an extreme form of bagging in which each model is trained on a single case and each parameter of the model is very strongly regularized by sharing it with the corresponding parameter in all the other models. **Referred here**
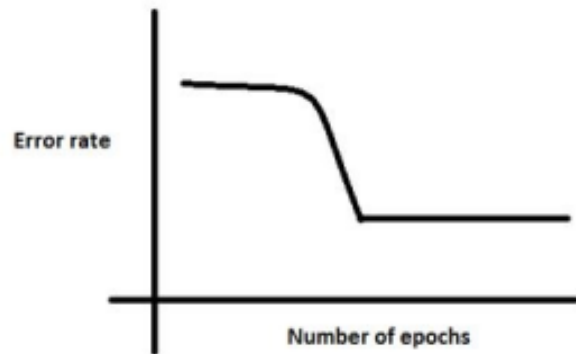
Q 12. Which of the following gives non-linearity to a neural network?

A. Stochastic Gradient DescentB. Rectified Linear Unit

C. Convolution function

Solution: (B)

Rectified Linear unit is a non-linear activation function.

Q13. In training a neural network, you notice that the loss does not decrease in the few starting epochs.

The reasons for this could be:

1. **The learning is rate is low**

2. **Regularization parameter is high**

3. **Stuck at local minima**

What according to you are the probable reasons?

A. 1 and 2

B. 2 and 3

C. 1 and 3

D. Any of these

Solution: (D)

The problem can occur due to any of the reasons mentioned.

Q14. Which of the following is true about model capacity (where model capacity means the ability of neural network to approximate complex functions) ?

B. As dropout ratio increases, model capacity increases

C. As learning rate increases, model capacity increases

D. None of these

Solution: (A)

Only option A is correct.

Q15. If you increase the number of hidden layers in a Multi Layer Perceptron, the classification error of test data always decreases. True or False?

A. TrueB. False

Solution: (B)

This is not always true. Overfitting may cause the error to increase.

Q16. You are building a neural network where it gets input from the previous layer as well as from itself.
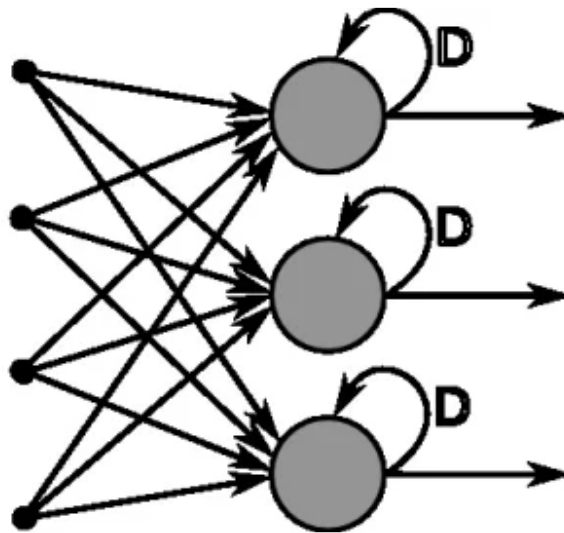
Which of the following architecture has feedback connections?

A. Recurrent Neural network

B. Convolutional Neural Network

C. Restricted Boltzmann Machine

Solution: (A)

Option A is correct.

Q17. What is the sequence of the following tasks in a perceptron?

1. **Initialize weights of perceptron randomly**

2. **Go to the next batch of dataset**

3. **If the prediction does not match the output, change the weights**

4. **For a sample input, compute an output**

A. 1, 2, 3, 4

B. 4, 3, 2, 1

C. 3, 1, 2, 4

D. 1, 4, 3, 2

Solution: (D)

Sequence D is correct.

Q18. Suppose that you have to minimize the [cost function](#)by changing the parameters. Which of the following technique could be used for this?
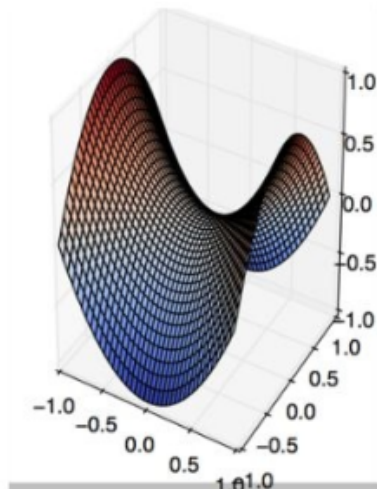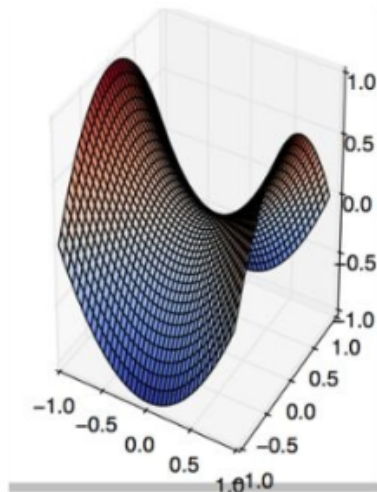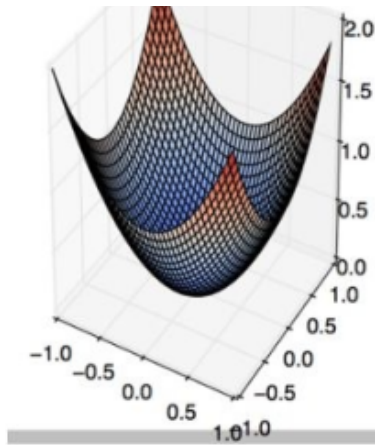
A. Exhaustive Search

B. Random Search

C. [Bayesian](#)Optimization

D. Any of these

Solution: (D)

Any of the above mentioned technique can be used to change parameters.

Q19. First Order Gradient descent would not work correctly (i.e. may get stuck) in which of the following graphs?

Solution: (B)

This is a classic example of saddle point problem of gradient descent.

Q20. The below graph shows the accuracy of a trained 3-layer convolutional neural network vs the number of parameters (i.e. number of feature kernels).
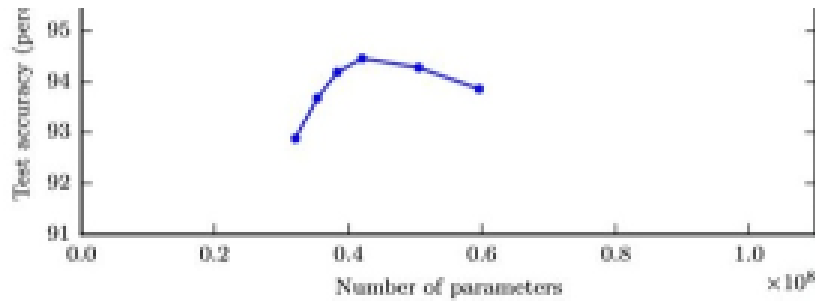
The trend suggests that as you increase the width of a neural network, the accuracy increases till a certain threshold value, and then starts decreasing.

What could be the possible reason for this decrease?

A. Even if number of kernels increase, only few of them are used for prediction

B. As the number of kernels increase, the predictive power of neural network decrease

C. As the number of kernels increase, they start to correlate with each other which in turn helps overfitting

D. None of these

Solution: (C)

As mentioned in option C, the possible reason could be kernel correlation.

Q21. Suppose we have one hidden layer neural network as shown above. The hidden layer in this network works as a dimensionality reductor. Now instead of using this hidden layer, we replace it with a dimensionality reduction technique such as PCA.
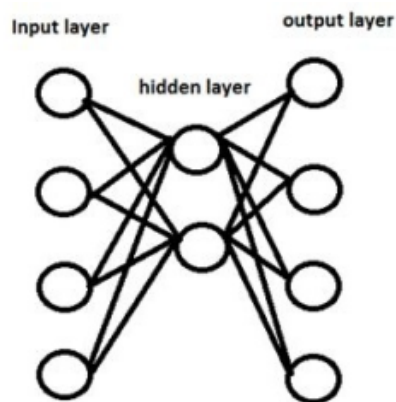
Would the network that uses a dimensionality reduction technique always give same output as network with hidden layer?

Solution: (B)

Because **PCA** works on correlated features, whereas hidden layers work on predictive capacity of features.

Q22. Can a neural network model the function (y=1/x)?

A. YesB. No

Solution: (A)

Option A is true, because activation function can be reciprocal function.

Q23. In which neural net architecture, does weight sharing occur?

A. Convolutional neural NetworkB. Recurrent Neural Network

C. Fully Connected Neural Network

D. Both A and B

Solution: (D)

Option D is correct.

Q24. Batch Normalization is helpful because

A. It normalizes (changes) all the input before sending it to the next layerB. It returns back the normalized mean and standard deviation of weights

C. It is a very efficient backpropagation technique

D. None of these

Solution: (A)

To read more about batch normalization, see refer **this video**

Q25. Instead of trying to achieve absolute zero error, we set a metric called bayes error which is the error we hope to achieve. What could be the reason for using bayes error?

A. Input variables may not contain complete information about the output variableB. System (that creates input-output mapping) may be stochastic

D. All the above

Solution: (D)

In reality achieving accurate prediction is a myth. So we should hope to achieve an "achievable result".

Q26. The number of neurons in the output layer should match the number of classes (Where the number of classes is greater than 2) in a supervised learning task. True or False?

A. TrueB. False

Solution: (B)

It depends on output encoding. If it is one-hot encoding, then its true. But you can have two outputs for four classes, and take the binary values as four classes(00,01,10,11).

Q27. In a neural network, which of the following techniques is used to deal with overfitting?

A. DropoutB. Regularization

C. Batch Normalization

D. All of these

Solution: (D)

All of the techniques can be used to deal with overfitting.

Q28. Y = ax^2 + bx + c (polynomial equation of degree 2)

Can this equation be represented by a neural network of single hidden layer with linear threshold?

A. Yes

B. No

Solution: (B)

The answer is no because having a linear threshold restricts your neural network and in simple terms, makes it a consequential linear transformation function.

Q29. What is a dead unit in a neural network?

B. A unit which does not respond completely to any of the training patterns

C. The unit which produces the biggest sum-squared error

D. None of these

Solution: (A)

Option A is correct.

Q30. Which of the following statement is the best description of early stopping?

A. Train the network until a local minimum in the error function is reached

B. Simulate the network on a test dataset after every epoch of training. Stop training when the generalization error starts to increase

C. Add a momentum term to the weight update in the Generalized Delta Rule, so that training converges more quickly

D. A faster version of backpropagation, such as the `Quickprop' algorithm
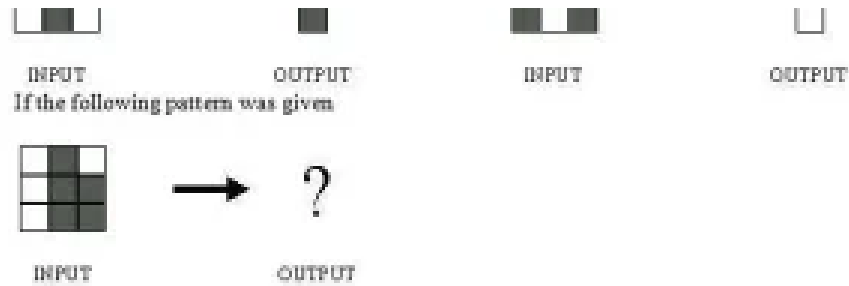
Solution: (B)

Option B is correct.

Q31. What if we use a learning rate that's too large?

A. Network will convergeB. Network will not converge

C. Can't Say

Solution: B

Option B is correct because the error rate would become erratic and explode.

Q32. The network shown in Figure 1 is trained to recognize the characters H and T as shown below:

INPUT   OUTPUT   INPUT   OUTPUT

If the following pattern was given

INPUT   OUTPUT

What would be the output of the network?

1.

2.

3.

4. Could be A or B depending on the weights of neural network

Solution: (D)

Without knowing what are the weights and biases of a neural network, we cannot comment on what output it would give.

Q33. Suppose a convolutional neural network is trained on ImageNet dataset (Object recognition dataset). This trained model is then given a completely white image as an input.The output probabilities for this input would be equal for all classes. True or False?

A. True

B. False

Solution: (B)

There would be some neurons which are do not activate for white pixels as input. So the classes wont be equal.

Q34. When pooling layer is added in a convolutional neural network, translation in-variance is preserved. True or False?

A. True

Solution: (A)

Translation invariance is induced when you use pooling.

Q35. Which gradient technique is more advantageous when the data is too big to handle in RAM simultaneously?

A. Full Batch Gradient Descent

B. Stochastic Gradient Descent

Solution: (B)

Option B is correct.

Q36. The graph represents gradient flow of a four-hidden layer neural network which is trained using sigmoid activation function per epoch of training. The neural network suffers with the vanishing gradient problem.
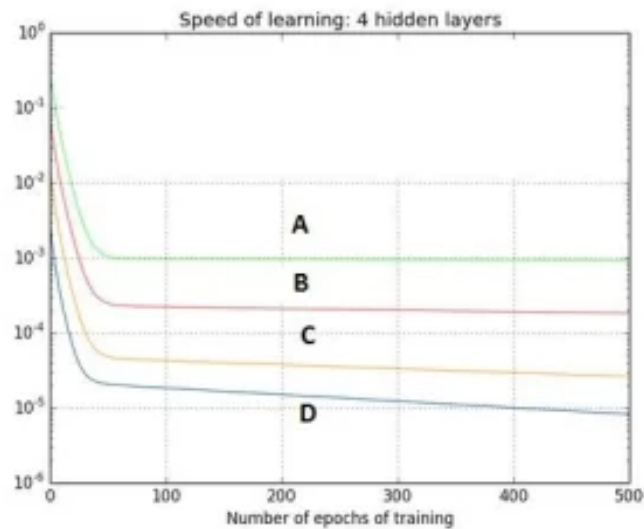
Which of the following statements is true?

A. Hidden layer 1 corresponds to D, Hidden layer 2 corresponds to C, Hidden layer 3 corresponds to B and Hidden layer 4 corresponds to A

B. Hidden layer 1 corresponds to A, Hidden layer 2 corresponds to B, Hidden layer 3 corresponds to C and Hidden layer 4 corresponds to D

Solution: (A)

Q37. For a classification task, instead of random weight initializations in a neural network, we set all the weights to zero. Which of the following statements is true?

A. There will not be any problem and the neural network will train properly

B. The neural network will train but all the neurons will end up recognizing the same thing

C. The neural network will not train as there is no net gradient change

D. None of these

Solution: (B)

Option B is correct.

Q38. There is a plateau at the start. This is happening because the neural network gets stuck at local minima before going on to global minima.
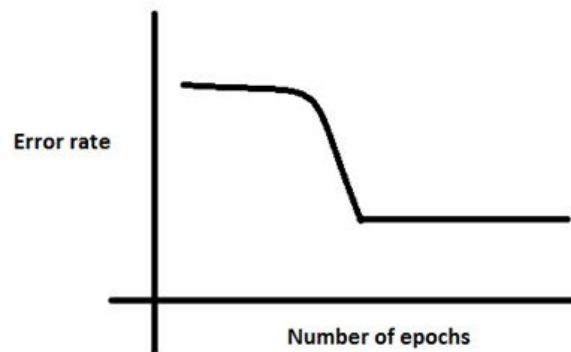
To avoid this, which of the following strategy should work?

A. Increase the number of parameters, as the network would not get stuck at local minima

B. Decrease the learning rate by 10 times at the start and then use momentum

C. Jitter the learning rate, i.e. change the learning rate for a few epochs

D. None of these

Solution: (C)

Option C can be used to take a neural network out of local minima in which it is stuck.

A. Multi Layer PerceptronB. Convolutional Neural Network

C. Recurrent Neural network

D. Perceptron

Solution: (B)

Convolutional Neural Network would be better suited for image related problems because of its inherent nature for taking into account changes in nearby locations of an image
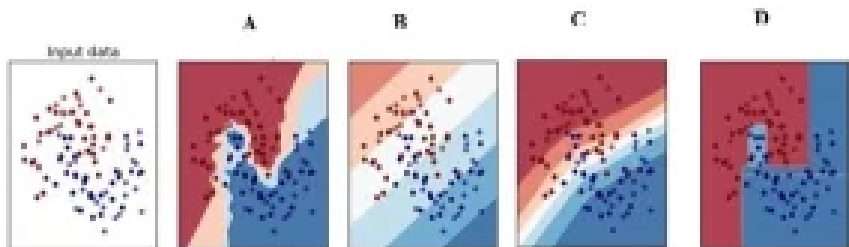
Q40. Suppose while training, you encounter this issue. The error suddenly increases after a couple of iterations.
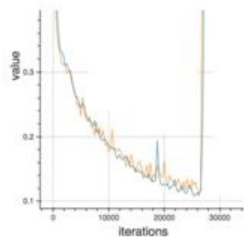
You determine that there must a problem with the data. You plot the data and find the insight that, original data is somewhat skewed and that may be causing the problem.
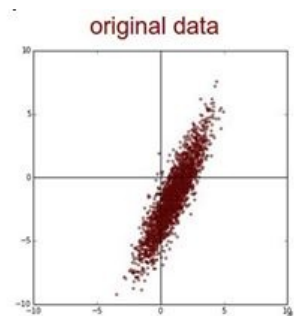
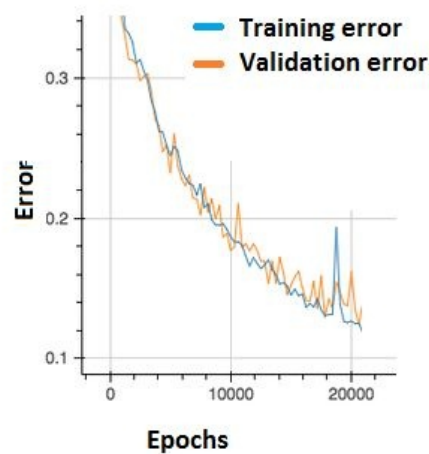**Q41. Which of the following is a decision boundary of Neural Network?**

A) B

C) D

D) C

E) All of these

Solution: (E)

A neural network is said to be a universal function approximator, so it can theoretically represent any decision boundary.

Q42. In the graph below, we observe that the error has many "ups and downs"

Should we be worried?

A. Yes, because this means there is a problem with the learning rate of neural network.

B. No, as long as there is a cumulative decrease in both training and validation error, we don't need to worry.

Solution: (B)

Option B is correct. In order to decrease these "ups and downs" try to increase the batch size.

Q43. What are the factors to select the depth of neural network?

1. Type of neural network (eg. MLP, **CNN**etc)

2. Input data

3. Computation power, i.e. Hardware capabilities and software capabilities

4. Learning Rate

5. The output function to map

B. 2, 3, 4, 5

C. 1, 3, 4, 5

D. All of these

Solution: (D)

All of the above factors are important to select the depth of neural network

Q44. Consider the scenario. The problem you are trying to solve has a small amount of data. Fortunately, you have a pre-trained neural network that was trained on a similar problem. Which of the following methodologies would you choose to make use of this pre-trained network?

A. Re-train the model for the new datasetB. Assess on every layer how the model performs and only select a few of them

C. Fine tune the last couple of layers only

D. Freeze all the layers except the last, re-train the last layer

Solution: (D)

If the dataset is mostly similar, the best method would be to train only the last layer, as previous all layers work as feature extractors.

Q45. Increase in size of a convolutional kernel would necessarily increase the performance of a convolutional network.

A. True

B. False

Solution: (B)

Conclusion

Increasing kernel size would not necessarily increase performance. This depends heavily on the dataset.

I hope you enjoyed taking the test and you found the solutions helpful. The test focused on conceptual knowledge of Deep Learning.

know by dropping your feedback in the comments section.

[Faizan Shaikh](#)

Faizan is a Data Science enthusiast and a Deep learning rookie. A recent Comp. Sc. undergrad, he aims to utilize his skills to push the boundaries of AI research.

Deep Learning   Intermediate   Interview Prep   Machine Learning   Skilltest

Free Courses

★ 4.7

Generative AI - A Way of Life

Explore Generative AI for beginners: create text and images, use top AI tools, learn practical skills, and ethics.

★ 4.5

Getting Started with Large Language Models

Master Large Language Models (LLMs) with this course, offering clear guidance in NLP and model training made simple.

★ 4.6

Building LLM Applications using Prompt Engineering

This free course guides you on building LLM apps, mastering prompt engineering, and developing chatbots with enterprise data.

Improving Real World RAG Systems: Key Challenges & Practical Solutions

Explore practical solutions, advanced retrieval strategies, and agentic RAG systems to improve context, relevance, and accuracy in AI-driven applications.

★ 4.7

Microsoft Excel: Formulas & Functions

Master MS Excel for data analysis with key formulas, functions, and LookUp tools in this comprehensive course.

Responses From Readers

What are your thoughts?...

Submit reply

For Q31. "What if we use a learning rate that's too large?", only three options (A,B,C) are available. But the answer is mentioned as E. I think options D, E are missing.

💬 1   Show 1 reply

Dai Zhongxiang

Thanks for the post. Can you provide more explanation on Q14? why would the model capacity decrease as the number of hidden layers increase? Intuitively, shouldn't it be the other way around? Thanks.
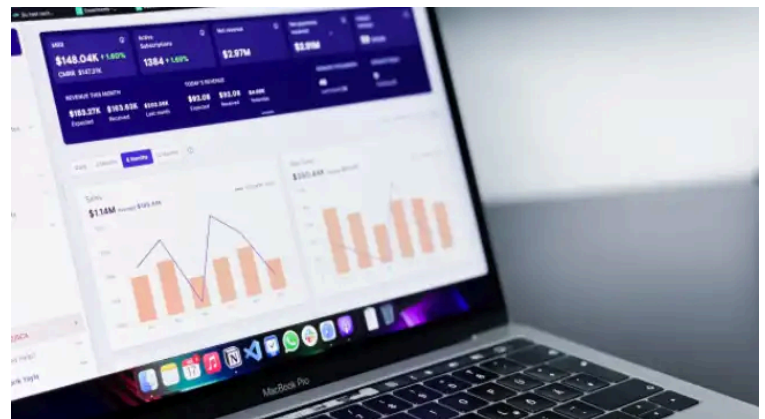
💬 1   Show 1 reply

Sukuya

Regarding Q1, how can a real neuron have multiple outputs ? When the signal

Write for us →

Write, captivate, and earn accolades and rewards for your work

- Reach a Global Audience
- Get Expert Feedback
- Build Your Brand & Audience

- Cash In on Your Knowledge
- Join a Thriving Community
- Level Up Your Data Science Game

🎲 Make This Article **Fun and Interactive** with Flash Cards and Quizzes!

Popular Categories

Generative AI | Prompt Engineering | Generative AI Application | News | Technical Guides | AI Tools | Interview Preparation | Research Papers | Success Stories | Quiz | Use Cases | Listicles

Generative AI Tools and Techniques

GANs | VAEs | Transformers | StyleGAN | Pix2Pix | Autoencoders | GPT | BERT | Word2Vec | LSTM | Attention Mechanisms | Diffusion Models | LLMs | SLMs | StyleGAN | Encoder Decoder Models | Prompt Engineering | LangChain | LlamaIndex | RAG | Fine-tuning | LangChain AI Agent | Multimodal Models | RNNs | DCGAN | ProGAN | Text-to-Image Models | DDPM | Document Question Answering | Imagen | T5 (Text-to-Text Transfer Transformer) | Seq2seq Models | WaveNet | Attention Is All You Need (Transformer Architecture)

Popular GenAI Models

Llama 3.1 | Llama 3 | Llama 2 | GPT 4o Mini | GPT 4o | GPT 3 | Claude 3 Haiku | Claude 3.5 Sonnet | Phi 3.5 | Phi 3 | Mistral Large 2 | Mistral NeMo | Mistral-7b | Gemini 1.5 Pro | Gemini Flash 1.5 | Bedrock | Vertex AI | DALL.E | Midjourney | Stable Diffusion

Data Science Tools and Techniques

Python | R | SQL | Jupyter Notebooks | TensorFlow | Scikit-learn | PyTorch | Tableau | Apache Spark | Matplotlib | Seaborn | Pandas | Hadoop | Docker | Git | Keras | Apache Kafka | AWS | NLP | Random Forest | Computer Vision | Data Visualization | Data Exploration | Big Data | Common Machine Learning Algorithms | Machine Learning

Make This Article **Fun and Interactive** with Flash Cards and Quizzes!

Events

Podcasts

Become a Mentor

Become an Instructor

Data Culture

AI Newsletter