QF632-2025-W7

Number of participants: 24

1. How many hidden layers does a three-layer ANN contain?

16 correct answers
out of 17 respondents

1	6%	1 vote
✅ 2	94%	16 votes
3	0%	0 votes

2. During training, you observe that if you increase the batch size from 32 to 256, the training loss decreases more smoothly but convergence slows down. You decide to run experiments with batch sizes $16,, 32,, 64,, 128,, 256$. This process is:

14 correct answers
out of 18 respondents

Parameter tuning, because batch size affects the mini-batch gradients that update weights.

6%

1 vote

Hyperparameter tuning, because batch size controls the training process but isn't a weight or bias.

78%

14 votes

Neither, because batch size only affects hardware utilization, not learning.

0%

0 votes

Both, since batch size influences gradient variance (parameter updates) and training speed (hyperparameter effect).

17%

3 votes

3. Which of the following statements is always true about a single-layer neural network (no hidden layers) with a linear activation function compared to ordinary least squares (OLS) linear regression?

11 correct answers out of 16 respondents

It can model interaction terms (e.g., products of input features) if weights are tuned appropriately.

31% 5 votes

Its loss surface (mean-squared error) is non-convex in the weights.

0% 0 votes

✓ It yields exactly the same fitted predictions as OLS linear regression when trained under MSE loss.

69% 11 votes

It can represent polynomial relationships between inputs and outputs without changing the architecture.
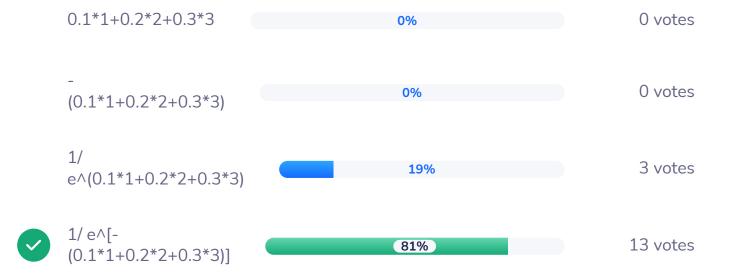
0% 0 votes

4. **Backpropagation means we take the input values and calculate the output of all layers until we reach the output layer**

12 correct answers
out of 14 respondents

True   14%   2 votes

✓ False   86%   12 votes

5. **What is the output value of node, with 3 links: H0: 1, w0: 0.1 | H1: 2, w1: 0.2 | H2: 3, w2: 0.3, & sigmoid activation?**

13 correct answers
out of 16 respondents

0.1*1+0.2*2+0.3*3   0%   0 votes

-
(0.1*1+0.2*2+0.3*3)   0%   0 votes

1/
e^(0.1*1+0.2*2+0.3*3)   19%   3 votes

✓ 1/ e^[-
(0.1*1+0.2*2+0.3*3)]   81%   13 votes
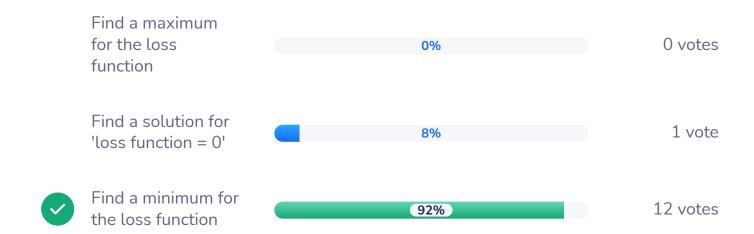
6. What happens when we increase the number of hidden layers in a neural network?

16 correct answers
out of 17 respondents

The model becomes more prone to underfitting.

0%

0 votes

The model becomes more prone to overfitting.

94%

16 votes

The model's training time decreases.

6%

1 vote

The model becomes less capable of learning complex patterns.

0%

0 votes

7. What is the gradient descent's objective?

12 correct answers
out of 13 respondents

Find a maximum for the loss function

0%

0 votes

Find a solution for 'loss function = 0'

8%

1 vote

Find a minimum for the loss function

92%

12 votes

8. Which application is not suitable for RNN?

3 correct answers
out of 15 respondents

Score audio recording of IELTS takers

27% — 4 votes

✓ Predicting default with 1000 customers' credit default data

20% — 3 votes

Predict stock movement using stock price of Apple over 10 years

33% — 5 votes

Translate a paragraph from English to Malay

20% — 3 votes

9. Which of the following can be a challenge when using SGD?

Choosing an appropriate learning rate.
0%
0 votes

Dealing with noisy gradients.
0%
0 votes

✓ Both a and b.
100%
12 votes

Neither a nor b.
0%
0 votes

10. **What is the main advantage of SGD over Batch Gradient Descent?**

8 correct answers out of 12 respondents

SGD is computationally less expensive.

67%

8 votes

SGD always finds the global minimum.

33%

4 votes

SGD uses more memory.

0%

0 votes

There is no advantage of SGD over Batch Gradient Descent.

0%

0 votes