

🔗 Google Discussions



Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)

📄 EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 1 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 1

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

Your company built a TensorFlow neural-network model with a large number of neurons and layers. The model fits well for the training data. However, when tested against new data, it performs poorly. What method can you employ to address this?

- A. Threading
- B. Serialization
- C. Dropout Methods
- D. Dimensionality Reduction

[Show Suggested Answer](#)

by [henriksoder24](#) at *Sept. 2, 2022, 2:46 p.m.*

Comments

Type your comment...

[Submit](#)

👤 [henriksoder24](#) Highly Voted 2 years, 8 months ago

Answer is C.

Bad performance of a model is either due to lack of relationship between dependent and independent variables used, or just overfit due to having used too many features and/or bad features.

A: Threading parallelisation can reduce training time, but if the selected features are the same then the resulting performance won't have changed

B: Serialization is only changing data into byte streams. This won't be useful.

C: This can show which features are bad. E.g. if it is one feature causing bad performance, then the dropout method will show it, so you can remove it from the model and retrain it.

D: This would become clear if the model did not fit the training data well. But the question says that the model fits the training data well, so D is not the answer.



   upvoted 28 times

  **su787** Most Recent 2 weeks ago

Selected Answer: C

Dropout method

   upvoted 1 times

  **monyu** 1 month, 4 weeks ago

Selected Answer: C

Correct answer is C.

A - Is not Threading because it is used to accelerate the training in order to reduce training time.

B - Is not Serialization because it transforms (serializes into bytes) the training data but does not increase or change the original nature.

D - Is not dimensionality reduction because the model fits the training data.


   upvoted 1 times

  **Ahamada** 2 months, 1 week ago

Selected Answer: C

Dropout methods is the solution here to resolve overfitting issue

   upvoted 1 times

  **onschamekh** 3 months, 2 weeks ago

Selected Answer: C

Dropout is a specific technique to prevent overfitting by randomly disabling a certain percentage of neurons during training. This helps the network avoid relying too heavily on a subset of neurons, thereby improving its ability to generalize to new data.

   upvoted 1 times

  **jithinlife** 4 months ago

Selected Answer: C

Can we expect similar questions like this in GCP exam as well?

   upvoted 1 times

  **SamuelTsch** 6 months, 2 weeks ago

Selected Answer: C

It occurs overfitting problem. A general idea is to simplify the model. A GENERALIZATION related method should be used.

   upvoted 2 times

  **rtcpost** 7 months, 1 week ago

Selected Answer: C

C. Dropout Methods

Dropout is a regularization technique commonly used in neural networks to prevent overfitting. It helps improve the generalization of the model by randomly setting a fraction of the neurons to zero during each training iteration, which prevents the network from relying too heavily on specific neurons. This, in turn, can lead to better performance on new, unseen data.

   upvoted 1 times

  **rocky48** 7 months, 1 week ago

Selected Answer: C

A: Threading parallelisation can reduce training time, but if the selected features are the same then the resulting performance won't have changed

B: Serialization is only changing data into byte streams. This won't be useful.

C: This can show which features are bad. E.g. if it is one feature causing bad performance, then the dropout method will show it, so you can remove it from the model and retrain it.

D: This would become clear if the model did not fit the training data well. But the question says that the model fits the training data well.

So, C is the answer.

   upvoted 1 times

  **trashbox** 12 months ago

Selected Answer: C

Dropout Methods are useful to prevent a TensorFlow model from overfitting


   upvoted 1 times

  **azmiozgen** 1 year, 9 months ago

Selected Answer: C

Answer is C. Dropout methods are used to mitigate overfitting. Hence, it is commonly used in training phase and it's beneficial for test-time performance.



   upvoted 1 times

  **dgteixeira** 1 year, 11 months ago

Selected Answer: C

Answer is C

   upvoted 1 times

  **AmmarFasih** 1 year, 11 months ago

Selected Answer: C

Dropout is a regularization technique commonly used in model training with TensorFlow and other deep learning frameworks. It is employed to prevent overfitting, a phenomenon where a model learns to perform well on the training data but fails to generalize well to new, unseen data.

   upvoted 1 times

  **IgnacioBL** 2 years, 1 month ago

Selected Answer: C

Answer is C

   upvoted 2 times

  **Morock** 2 years, 2 months ago

Selected Answer: C

Dropout is a regularization method to remove random selection of fixed number of unit in a neural network layer. So pick C for this question.

   upvoted 1 times

  **enghabeth** 2 years, 2 months ago

Selected Answer: C

because it is a regularization technique for reducing overfitting in neural networks by preventing complex co-adaptations on training data.

   upvoted 1 times

  **samdhimal** 2 years, 3 months ago

C. Dropout Methods

Dropout is a regularization technique that can be used to prevent overfitting of the model to the training data. It works by randomly dropping out a certain percentage of neurons during training, which helps to reduce the complexity of the model and prevent it from memorizing the training data. This can improve the model's ability to generalize to new data and reduce the risk of poor performance when tested against new data.

   upvoted 4 times

  **samdhimal** 2 years, 3 months ago

A. Threading: it's not a method to address overfitting, it's a technique to improve the performance of the model by parallelizing the computations using multiple threads.

B. Serialization: it's a technique to save the model's architecture and trained parameters to a file, it's helpful when you want to reuse the model later, but it doesn't address overfitting problem.

D. Dimensionality Reduction: it's a technique that can be used to reduce the number of features in the data, it's helpful when the data contains redundant or irrelevant features, but it doesn't address overfitting problem directly.

   upvoted 4 times

[Load full discussion...](#)

Platform

- > Home
- > Examtopics PRO
- > All Exams
- > Training Courses

