← **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 191 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 191

Topic #: 1

[All Professional Data Engineer Questions]

You are developing a new deep learning model that predicts a customer's likelihood to buy on your ecommerce site. After running an evaluation of the model against both the original training data and new test data, you find that your model is overfitting the data. You want to improve the accuracy of the model when predicting new data. What should you do?

  A. Increase the size of the training dataset, and increase the number of input features.

  B. Increase the size of the training dataset, and decrease the number of input features.

  C. Reduce the size of the training dataset, and increase the number of input features.

  D. Reduce the size of the training dataset, and decrease the number of input features.

Show Suggested Answer

by 👤 **ducc** at *Sept. 3, 2022, 3:46 a.m.*

Comments

Type your comment...

Submit

☐ 👤 **John_Pongthorn** [Highly Voted 👍] 2 years, 1 month ago

Selected Answer: B

There 2 parts and they are relevant to each other
1. Overfit is fixed by decreasing the number of input features (select only essential features)
2. Accuracy is improved by increasing the amount of training data examples.

👍 ↩ 🏳 upvoted 11 times

    ⊟ 👤 **John_Pongthorn** 2 years, 1 month ago

    https://docs.aws.amazon.com/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html

    👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **Matt_108** `Most Recent ⊘` 9 months, 3 weeks ago

`Selected Answer: B`

Option B, the model learned to listen to too much stuff/noise. We need to reduce it, by decreasing the number of input feature, and we need to give the model more data, by increasing the amount of training data

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **NeoNitin** 1 year, 3 months ago

Increase the size of the training dataset: By adding more diverse examples of customers and their buying behavior to the training data, the model will have a broader understanding of different scenarios and be better equipped to generalize to new customers.

Increase the number of input features: Providing the model with more relevant information about customers can help it identify meaningful patterns and make better predictions. These input features could include things like the customer's age, past purchase history, browsing behavior, or any other relevant data that might impact their buying likelihood.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **vaga1** 1 year, 5 months ago

`Selected Answer: B`

A. can be a solution for a specific case, but it is not the academic answer as we do not know the quantity and proportion between them of n and k added. More records and more variables together can lead to even more overfitting due also to the curse of dimensionality. Adding a variable is much more impactful than records.
B. just more records can lead to a more robust estimation and fewer variables certainly lead to at most the same estimation, but potentially reduce the fit on the training set.
C. reduce n in favor of k is never a choice. it is against logic and it will lead to more overfitting.
D. decrease both will reduce overfitting for sure but at the price of losing robustness on the model predictive power

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **AzureDP900** 1 year, 10 months ago

B. Increase the size of the training dataset, and decrease the number of input features.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **pluiedust** 2 years, 1 month ago

`Selected Answer: B`

B is correct

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **TNT87** 2 years, 1 month ago

Answer B
https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and-performance-estimates/

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **HarshKothari21** 2 years, 1 month ago

`Selected Answer: B`

Option B
Feature selection is the one the ways to resolve overfitting. Which means reducing the features
when the size of the training data is small, then the network tends to have greater control over the training data. so increasing the size of data would help.

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **YorelNation** 2 years, 2 months ago

`Selected Answer: B`

Best option is not mentioned: generalize you neural net by decreasing the complexity of it's structure.

A part from that I guess you could remove some features and increase the size of the training dataset ==> B

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **AWSandeep** 2 years, 2 months ago

`Selected Answer: B`

B. Increase the size of the training dataset, and decrease the number of input features.

Sorry, B is right. Read through extensive best-practices on ML.

upvoted 1 times

☐ 👤 **ducc** 2 years, 2 months ago

Selected Answer: D

D is correct

👍 ↩ ⚑ upvoted 1 times

☐ 👤 **AWSandeep** 2 years, 2 months ago

D. Reduce the size of the training dataset, and decrease the number of input features.
Reveal Solution

👍 ↩ ⚑ upvoted 1 times

☐ 👤 **ducc** 2 years, 2 months ago

Selected Answer: B

B. Increase the size of the training dataset, and decrease the number of input features.

👍 ↩ ⚑ upvoted 1 times