

 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 51 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 51

Topic #: 1

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

You are training a spam classifier. You notice that you are overfitting the training data. Which three actions can you take to resolve this problem? (Choose three.)

- A. Get more training examples
- B. Reduce the number of training examples
- C. Use a smaller set of features
- D. Use a larger set of features
- E. Increase the regularization parameters
- F. Decrease the regularization parameters

[Show Suggested Answer](#)

by  [madhu1171](#) at March 13, 2020, 10:28 a.m.

Comments

Type your comment...

[Submit](#)

— **maund177** **Highly Voted** 4 years, 7 months ago

it should be ACE

👍 ↩ 🚩 upvoted 68 times

🗨️ **[Removed]** **Highly Voted** 4 years, 7 months ago

Should be ACE

👍 ↩ 🚩 upvoted 19 times

🗨️ **[Removed]** 4 years, 7 months ago

prevent overfitting: less variables, regularisation, early ending on the training

👍 ↩ 🚩 upvoted 14 times

🗨️ **monyu** **Most Recent** 1 month, 2 weeks ago

Selected Answer: ACE

A. Because getting more training samples reduces significantly the risk of overfitting since the algorithm can learn from a more general dataset.

C. Introducing lots of features increases the risk to introducing irrelevant information, driving the model to avoid focusing on the truly important patterns.

E. Because regularization increases the penalty term to the loss function, which discourages complex models with large coefficients avoiding overfitting.

👍 ↩ 🚩 upvoted 1 times

🗨️ **TVH_Data_Engineer** 10 months, 3 weeks ago

Selected Answer: ACE

To address the problem of overfitting in training a spam classifier, you should consider the following three actions:

A. Get more training examples:

Why: More training examples can help the model generalize better to unseen data. A larger dataset typically reduces the chance of overfitting, as the model has more varied examples to learn from.

C. Use a smaller set of features:

Why: Reducing the number of features can help prevent the model from learning noise in the data. Overfitting often occurs when the model is too complex for the amount of data available, and having too many features can contribute to this complexity.

E. Increase the regularization parameters:

Why: Regularization techniques (like L1 or L2 regularization) add a penalty to the model for complexity. Increasing the regularization parameter will strengthen this penalty, encouraging the model to be simpler and thus reducing overfitting.

👍 ↩ 🚩 upvoted 5 times

🗨️ **Mathew106** 1 year, 3 months ago

Selected Answer: ACE

100% ACE

We need more data because less data induces overfitting. We need less features to make the problem simpler to learn and not promote learning a very complex function for thousands of features that might not apply to the test data. We also need to use regularization to keep the weights constrained.

👍 ↩ 🚩 upvoted 2 times

🗨️ **theseawillclaim** 1 year, 3 months ago

Selected Answer: ACE

Definitely ACE.

More training data and less variables can prevent the model from being too picky or specific.

👍 ↩ 🚩 upvoted 1 times

🗨️ **jin0** 1 year, 8 months ago

? why A is answer? even though 'more training example' not 'more dataset example'. I understand that there is dataset same and there is only change the size of training examples size. in this case there are valid and test example should be reduced. isn't it?

👍 ↩ 🚩 upvoted 1 times

🗨️ **desertlotus1211** 1 year, 9 months ago



Collect more training data: This will help the model generalize better and reduce overfitting.

Use regularization techniques: Techniques such as L1 and L2 regularization can be applied to the model's weights to prevent them from becoming too large and causing overfitting.

Use early stopping: This involves monitoring the performance of the model on a validation set during training, and stopping

the training when the performance on the validation set starts to degrade. This helps to prevent the model from becoming too complex and overfitting the training data.

   upvoted 1 times

  **desertlotus1211** 1 year, 9 months ago

Regularization is a technique that penalizes the coefficient. In an overfit model, the coefficients are generally inflated. Thus, Regularization adds penalties to the parameters and avoids them weigh heavily.

A & C are correct... the third one --- not sure on

   upvoted 1 times

  **RoshanAshraf** 1 year, 10 months ago

Selected Answer: ACE

A -The training data is causing the overfitting for the testing data, so addition of training data will solve this.
C - Larger sets will cause overfitting, so we have to use smaller sets or reduce features
E - Increase the regularization is a method for solving the Overfitting model

   upvoted 1 times

  **AzureDP900** 1 year, 10 months ago

Answers are;


A. Get more training examples
C. Use a smaller set of features
E. Increase the regularization parameters

Prevent overfitting: less variables, regularisation, early ending on the training

Reference:

<https://cloud.google.com/bigquery-ml/docs/preventing-overfitting>

   upvoted 3 times

  **DGames** 1 year, 10 months ago

Selected Answer: ADE

Answer ADE

   upvoted 1 times

  **MisuLava** 2 years ago

Selected Answer: ACE

100% sure ACE

<https://elitedatascience.com/overfitting-in-machine-learning>

   upvoted 1 times

  **MisuLava** 2 years, 2 months ago

Answer is : ACE

<https://www.ibm.com/cloud/learn/overfitting#:~:text=Overfitting%20is%20a%20concept%20in,unseen%20data%2C%20defeating%20its%20purpose.>

   upvoted 1 times

  **Noahz110** 2 years, 2 months ago

Selected Answer: ACE

im vote for ACE

   upvoted 1 times

  **Dip1994** 2 years, 3 months ago

It should be ACE

   upvoted 1 times

  **sraakesh95** 2 years, 9 months ago

Selected Answer: ACE

@medeis_jar

   upvoted 1 times

  **medeis_jar** 2 years, 10 months ago

Selected Answer: ACE

As MaxNRG wrote:

The tools to prevent overfitting: less variables, regularization, early ending on the training.

- Adding more training data will increase the complexity of the training set and help with the variance problem.
- Reducing the feature set will ameliorate the overfitting and help with the variance problem.

Increasing the regularization parameter will reduce overfitting and help with the variance problem.

- increasing the regularization parameter will reduce overfitting and help with the variance problem.

👍 ↩️ 🚩 upvoted 4 times

[Load full discussion...](#)



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics