

Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 78 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 78

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You are creating a deep neural network classification model using a dataset with categorical input values. Certain columns have a cardinality greater than 10,000 unique values. How should you encode these categorical values as input into the model?

- A. Convert each categorical value into an integer value.
- B. Convert the categorical string data to one-hot hash buckets.
- C. Map the categorical variables into a vector of boolean values.
- D. Convert each categorical value into a run-length encoded string.

Show Suggested Answer

by [LearnSodas](#) at Dec. 11, 2022, 5 p.m.

Comments

Type your comment...

Submit

[PhilipKoku](#) 5 months ago

Selected Answer: B

B) Hash buckets



   upvoted 1 times

  **etienne0** 8 months, 1 week ago

Selected Answer: A

went with A

   upvoted 1 times

  **M25** 1 year, 6 months ago

Selected Answer: B

Went with B

   upvoted 2 times

  **CloudKida** 1 year, 6 months ago

Selected Answer: B

<https://cloud.google.com/ai-platform/training/docs/algorithms/wide-and-deep>

If the column is categorical with high cardinality, then the column is treated with hashing, where the number of hash buckets equals to the square root of the number of unique values in the column.

   upvoted 4 times

  **JamesDoe** 1 year, 7 months ago

Selected Answer: B

B.

The other options solves nada.

   upvoted 1 times

  **enghabeth** 1 year, 9 months ago

Selected Answer: B

<https://towardsdatascience.com/getting-deeper-into-categorical-encodings-for-machine-learning-2312acd347c8>

When you have millions uniques values try to do: Hash Encoding

   upvoted 1 times

  **John_Pongthorn** 1 year, 9 months ago

Selected Answer: B

B unconditoinally

<https://cloud.google.com/ai-platform/training/docs/algorithms/xgboost#analysis>

If the column is categorical with high cardinality, then the column is treated with hashing, where the number of hash buckets equals to the square root of the number of unique values in the column.

A categorical column is considered to have high cardinality if the number of unique values is greater than the square root of the number of rows in the dataset.



   upvoted 2 times

  **MithunDesai** 1 year, 10 months ago

Selected Answer: C

I think C as it has 10000 categorical values

   upvoted 2 times

  **hiromi** 1 year, 10 months ago

Selected Answer: B

I think B is correct

Ref.:"

- <https://cloud.google.com/ai-platform/training/docs/algorithms/xgboost>


- <https://stackoverflow.com/questions/26473233/in-preprocessing-data-with-high-cardinality-do-you-hash-first-or-one-hot-encode>

   upvoted 4 times

  **hiromi** 1 year, 10 months ago

- <https://cloud.google.com/ai-platform/training/docs/algorithms/xgboost#analysis>



   upvoted 1 times

  **mil_spyro** 1 year, 10 months ago

Selected Answer: B



Answer is B. When cardinality of the categorical column is very large best choice is binary encoding however it not here hence one-hot hash option.

   upvoted 1 times

  **mil_spyro** 1 year, 10 months ago

<https://www.analyticsvidhya.com/blog/2020/08/types-of-categorical-data-encoding/>



   upvoted 1 times

  **JeanEl** 1 year, 10 months ago

Selected Answer: B

Ans : B



   upvoted 1 times

  **seifou** 1 year, 10 months ago

Selected Answer: B

B is correct

   upvoted 1 times

  **ares81** 1 year, 10 months ago

It should be B

   upvoted 1 times

  **LearnSodas** 1 year, 11 months ago

Selected Answer: A

Answer A since with 10.000 unique values one-hot shouldn't be a good solution

<https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python/>

   upvoted 3 times

  **etienne0** 8 months, 1 week ago

I agree with A

   upvoted 1 times

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)

EXAMTOPICS

We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.