41. best-practices-improve-input-output-performance-tensorflow-model-structured-dataset

Question

You are training a TensorFlow model on a structured dataset with 100 billion records stored in several CSV files.

You need to improve the input/output execution performance.

What should you do?

Answers

○ A. Load the data into BigQuery, and read the data from BigQuery.

○ B. Load the data into Cloud Bigtable, and read the data from Bigtable.

○ C. Convert the CSV files into shards of TFRecords, and store the data in Cloud Storage.

○ D. Convert the CSV files into shards of TFRecords, and store the data in the Hadoop Distributed File System (HDFS).

Show Answer

Prev Question   Next Question

42. best-method-deploy-tensorflow-model-daily-usage

Using TensorFlow Model for Daily Data Aggregation | ML Engineer Guide

Prev Question   Next Question   Go to: ▾

Question

As the lead ML Engineer for your company, you are responsible for building ML models to digitize scanned customer forms.

You have developed a TensorFlow model that converts the scanned images into text and stores them in Cloud Storage.

You need to use your ML model on the aggregated data collected at the end of each day with minimal manual intervention.

What should you do?

Answers

○ A. Use the batch prediction functionality of AI Platform.

○ B. Create a serving pipeline in Compute Engine for prediction.

○ C. Use Cloud Functions for prediction each time a new data point is ingested.

○ D. Deploy the model on AI Platform and create a version of it for online inference.

Show Answer

43.

Question

You recently joined an enterprise-scale company that has thousands of datasets.

You know that there are accurate descriptions for each table in BigQuery, and you are searching for the proper BigQuery table to use for a model you are building on AI Platform.

How should you find the data that you need?

Answers

○ A. Use Data Catalog to search the BigQuery datasets by using keywords in the table description.

○ B. Tag each of your model and version resources on AI Platform with the name of the BigQuery table that was used for training.

○ C. Maintain a lookup table in BigQuery that maps the table descriptions to the table I.

○ D. Query the lookup table to find the correct table ID for the data that you need.

○ E. Execute a query in BigQuery to retrieve all the existing table names in your project using the INFORMATION_SCHEMA metadata tables that are native to BigQuery. Use the result o find the table that you need.

Show Answer

44. best-approach-address-overfitting

Next Steps for Identifying and Fixing the Problem in Time Series Data Classification with High AUC ROC Value

Question

You started working on a classification problem with time series data and achieved an area under the receiver operating characteristic curve (AUC ROC) value of 99% for training data after just a few experiments.

You haven't explored using any sophisticated algorithms or spent any time on hyperparameter tuning.

What should your next step be to identify and fix the problem?

Answers

○ A. Address the model overfitting by using a less complex algorithm.

○ B. Address data leakage by applying nested cross-validation during model training.

○ C. Address data leakage by removing features highly correlated with the target value.

○ D. Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.

Show Answer

45. prediction-pipeline-configuration

Question

You work for an online travel agency that also sells advertising placements on its website to other companies.

You have been asked to predict the most relevant web banner that a user should see next.

Security is important to your company.

The model latency requirements are 300ms@p99, the inventory is thousands of web banners, and your exploratory analysis has shown that navigation context is a good predictor.

You want to Implement the simplest solution.

How should you configure the prediction pipeline?

Answers

○ A. Embed the client on the website, and then deploy the model on AI Platform Prediction.

○ B. Embed the client on the website, deploy the gateway on App Engine, and then deploy the model on AI Platform Prediction.

○ C. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Cloud Bigtable for writing and for reading the user`s navigation context, and then deploy the model on AI Platform Prediction.

○ D. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Memorystore for writing and for reading the user`s navigation context, and then deploy the model on Google Kubernetes Engine.

Show Answer

Prev Question   Next Question

46. the-best-environment-to-train-machine-learning-model

Question

Your team is building a convolutional neural network (CNN)-based architecture from scratch.

The preliminary experiments running on your on-premises CPU-only infrastructure were encouraging, but have slow convergence.

You have been asked to speed up model training to reduce time-to-market.

You want to experiment with virtual machines (VMs) on Google Cloud to leverage more powerful hardware.

Your code does not include any manual device placement and has not been wrapped in Estimator model-level abstraction.

Which environment should you train your model on?

Answers

○ A. AVM on Compute Engine and 1 TPU with all dependencies installed manually.

○ B. AVM on Compute Engine and 8 GPUs with all dependencies installed manually.

○ C. A Deep Learning VM with an n1-standard-2 machine and 1 GPU with all libraries pre-installed.

○ D. A Deep Learning VM with more powerful CPU e2-highcpu-16 machines with all libraries pre-installed.

Show Answer

Prev Question   Next Question

47. organizing-jobs-models-versions-ai-platform

Question

You work on a growing team of more than 50 data scientists who all use AI Platform.

You are designing a strategy to organize your jobs, models, and versions in a clean and scalable way.

Which strategy should you choose?

Answers

○ A. Set up restrictive IAM permissions on the AI Platform notebooks so that only a single user or group can access a given instance.

○ B. Separate each data scientist`s work into a different project to ensure that the jobs, models, and versions created by each data scientist are accessible only to that user.

○ C. Use labels to organize resources into descriptive categories. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources.

○ D. Set up a BigQuery sink for Cloud Logging logs that is appropriately filtered to capture information about AI Platform resource usage. In BigQuery, create a SQL view that maps users to the resources they are using.

Show Answer

Prev Question Next Question

48. semantic-image-segmentation-training-error

Question

You are training a deep learning model for semantic image segmentation with reduced training time.

While using a Deep Learning VM Image, you receive the following error: The resource 'projects/deeplearning-platform/zones/europe-west4-c/acceleratorTypes/nvidia-tesla-k80' was not found.

What should you do?

Answers

○ A. Ensure that you have GPU quota in the selected region.

○ B. Ensure that the required GPU is available in the selected region.

○ C. Ensure that you have preemptible GPU quota in the selected region.

○ D. Ensure that the selected GPU has enough GPU memory for the workload.

Show Answer

Prev Question Next Question

49.

Question

Your team is working on an NLP research project to predict political affiliation of authors based on articles they have written.

You have a large training dataset that is structured like this:

```
AuthorA:Political Party A
    TextA1: [SentenceA11, SentenceA12, SentenceA13, ...]
    TextA2: [SentenceA21, SentenceA22, SentenceA23, ...]
    ...
AuthorB:Political Party B
    TextB1: [SentenceB11, SentenceB12, SentenceB13, ...]
    TextB2: [SentenceB21, SentenceB22, SentenceB23, ...]
    ...
AuthorC:Political Party B
    TextC1: [SentenceC11, SentenceC12, SentenceC13, ...]
    TextC2: [SentenceC21, SentenceC22, SentenceC23, ...]
    ...
AuthorD:Political Party A
    TextD1: [SentenceD11, SentenceD12, SentenceD13, ...]
    TextD2: [SentenceD21, SentenceD22, SentenceD23, ...]
    ...
...
```

You followed the standard 80%-10%-10% data distribution across the training, testing, and evaluation subsets.

How should you distribute the training examples across the train-test-eval subsets while maintaining the 80-10-10 proportion?

Answers

○ A. Distribute texts randomly across the train-test-eval subsets: Train set: [TextA1, TextB2, ...] Test set: [TextA2, TextC1, TextD2, ...] Eval set: [TextB1, TextC2, TextD1, ...]

○ B. Distribute authors randomly across the train-test-eval subsets: (*) Train set: [TextA1, TextA2, TextD1, TextD2, ...] Test set: [TextB1, TextB2, ...] Eval set: [TexC1,TextC2 ...]

○ C. Distribute sentences randomly across the train-test-eval subsets: Train set: [SentenceA11, SentenceA21, SentenceB11, SentenceB21, SentenceC11, SentenceD21 ...] Test set: [SentenceA12, SentenceA22, SentenceB12, SentenceC22, SentenceC12, SentenceD22 ...] Eval set: [SentenceA13, SentenceA23, SentenceB13, SentenceC23, SentenceC13, SentenceD31 ...]

○ D. Distribute paragraphs of texts (i.e., chunks of consecutive sentences) across the train-test-eval subsets: Train set: [SentenceA11, SentenceA12, SentenceD11, SentenceD12 ...] Test set: [SentenceA13, SentenceB13, SentenceB21, SentenceD23, SentenceC12, SentenceD13 ...] Eval set: [SentenceA11, SentenceA22, SentenceB13, SentenceD22, SentenceC23, SentenceD11 ...]

Show Answer

50. best-classifier-ml-solution-google-cloud

Building a Classifier with TensorFlow and Kubeflow Pipelines

Question

Your team has been tasked with creating an ML solution in Google Cloud to classify support requests for one of your platforms.

You analyzed the requirements and decided to use TensorFlow to build the classifier so that you have full control of the model's code, serving, and deployment.

You will use Kubeflow pipelines for the ML platform.

To save time, you want to build on existing resources and use managed services instead of building a completely new model.

How should you build the classifier?

Answers

○ A. Use the Natural Language API to classify support requests.

○ B. Use AutoML Natural Language to build the support requests classifier.

○ C. Use an established text classification model on AI Platform to perform transfer learning.

○ D. Use an established text classification model on AI Platform as-is to classify support requests.

Show Answer

51. production-readiness-ml-components

Question

You recently joined a machine learning team that will soon release a new project.

As a lead on the project, you are asked to determine the production readiness of the ML components.

The team has already tested features and data, model development, and infrastructure.

Which additional readiness check should you recommend to the team?

Answers

○ A. Ensure that training is reproducible.

○ B. Ensure that all hyperparameters are tuned.

○ C. Ensure that model performance is monitored.

○ D. Ensure that feature expectations are captured in the schema.

Show Answer

Prev Question   Next Question

52. fraud-detection-model-optimization-objective

Question

You work for a credit card company and have been asked to create a custom fraud detection model based on historical data using AutoML Tables.

You need to prioritize detection of fraudulent transactions while minimizing false positives.

Which optimization objective should you use when training the model?

Answers

○ A. An optimization objective that minimizes Log loss

○ B. An optimization objective that maximizes the Precision at a Recall value of 0.50

○ C. An optimization objective that maximizes the area under the precision-recall curve (AUC PR) value

○ D. An optimization objective that maximizes the area under the receiver operating characteristic curve (AUC ROC) value.

Show Answer

Prev Question   Next Question

53. the-best-model-to-predict-popular-videos

Question

Your company manages a video sharing website where users can watch and upload videos.

You need to create an ML model to predict which newly uploaded videos will be the most popular so that those videos can be prioritized on your company's website.

Which result should you use to determine whether the model is successful?

Answers

○ A. The model predicts videos as popular if the user who uploads them has over 10,000 likes.

○ B. The model predicts 97.5% of the most popular clickbait videos measured by number of clicks.

○ C. The model predicts 95% of the most popular videos measured by watch time within 30 days of being uploaded.

○ D. The Pearson correlation coefficient between the log-transformed number of views after 7 days and 30 days after publication is equal to 0.

Show Answer

Prev Question   Next Question

54. neural-network-project-data-normalization

Question

You are working on a Neural Network-based project.

The dataset provided to you has columns with different ranges.

While preparing the data for model training, you discover that gradient optimization is having difficulty moving weights to a good solution.

What should you do?

Answers

○ A. Use feature construction to combine the strongest features.

○ B. Use the representation transformation (normalization) technique.

○ C. Improve the data cleaning step by removing features with missing values.

○ D. Change the partitioning step to reduce the dimension of the test set and have a larger training set.

Show Answer

Prev Question | Next Question

55. data-science-experiment-tracking-google-cloud-platform

Question

Your data science team needs to rapidly experiment with various features, model architectures, and hyperparameters.

They need to track the accuracy metrics for various experiments and use an API to query the metrics over time.

What should they use to track and report their experiments while minimizing manual effort?

Answers

○ A. Use Kubeflow Pipelines to execute the experiments. Export the metrics file, and query the results using the Kubeflow Pipelines API.

○ B. Use AI Platform Training to execute the experiments. Write the accuracy metrics to BigQuery, and query the results using the BigQuery API.

○ C. Use AI Platform Training to execute the experiments. Write the accuracy metrics to Cloud Monitoring, and query the results using the Monitoring API.

○ D. Use AI Platform Notebooks to execute the experiments. Collect the results in a shared Google Sheets file, and query the results using the Google Sheets API.

Show Answer

Prev Question | Next Question

56. best-data-transformation-strategy-fraud-detection

Question

You work for a bank and are building a random forest model for fraud detection.

You have a dataset that includes transactions, of which 1% are identified as fraudulent.

Which data transformation strategy would likely improve the performance of your classifier?

Answers

○ A. Write your data in TFRecords.

○ B. Z-normalize all the numeric features.

○ C. Oversample the fraudulent transaction 10 times.

○ D. Use one-hot encoding on all categorical features.

Show Answer

Prev Question