

Google

Exam Questions Professional-Machine-Learning-Engineer

Google Professional Machine Learning Engineer



NEW QUESTION 1

You want to rebuild your ML pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over 12 hours to run. To speed up development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting the speed and processing requirements?

- A. Use Data Fusion's GUI to build the transformation pipelines, and then write the data into BigQuery
- B. Convert your PySpark into SparkSQL queries to transform the data and then run your pipeline on Dataproc to write the data into BigQuery.
- C. Ingest your data into Cloud SQL convert your PySpark commands into SQL queries to transform the data, and then use federated queries from BigQuery for machine learning
- D. Ingest your data into BigQuery using BigQuery Load, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table

Answer: B

NEW QUESTION 2

Your organization wants to make its internal shuttle service route more efficient. The shuttles currently stop at all pick-up points across the city every 30 minutes between 7 am and 10 am. The development team has already built an application on Google Kubernetes Engine that requires users to confirm their presence and shuttle station one day in advance. What approach should you take?

- A. 1. Build a tree-based regression model that predicts how many passengers will be picked up at each shuttle station.* 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the prediction.
- B. 1. Build a tree-based classification model that predicts whether the shuttle should pick up passengers at each shuttle station.* 2. Dispatch an available shuttle and provide the map with the required stops based on the prediction
- C. 1. Define the optimal route as the shortest route that passes by all shuttle stations with confirmed attendance at the given time under capacity constraints.* 2 Dispatch an appropriately sized shuttle and indicate the required stops on the map
- D. 1. Build a reinforcement learning model with tree-based classification models that predict the presence of passengers at shuttle stops as agents and a reward function around a distance-based metric* 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the simulated outcome.

Answer: D

NEW QUESTION 3

You need to design a customized deep neural network in Keras that will predict customer purchases based on their purchase history. You want to explore model performance using multiple model architectures, store training data, and be able to compare the evaluation metrics in the same dashboard. What should you do?

- A. Create multiple models using AutoML Tables
- B. Automate multiple training runs using Cloud Composer
- C. Run multiple training jobs on AI Platform with similar job names
- D. Create an experiment in Kubeflow Pipelines to organize multiple runs

Answer: C

NEW QUESTION 4

You are going to train a DNN regression model with Keras APIs using this code:

```
model = tf.keras.Sequential()
model.add(tf.keras.layers.Dense (
    256,
    use_bias=True,
    activation='relu',
    kernel_initializer=None,
    kernel_regularizer=None,
    input_shape=(500,)))
model.add(tf.keras.layers.Dropout (rate=0.25))
model.add(tf.keras.layers.Dense (
    128, use_bias=True,
    activation='relu',
    kernel_initializer='uniform',
    kernel_regularizer='l2'))
model.add(tf.keras.layers.Dropout (rate=0.25))
model.add(tf.keras.layers.Dense (
    2, use_bias=False,
    activation='softmax'))
model.compile(loss='mse')
```

How many trainable weights does your model have? (The arithmetic below is correct.)

- A. $501 \times 256 + 257 \times 128 + 2 = 161154$
- B. $500 \times 256 + 256 \times 128 + 128 \times 2 = 161024$
- C. $501 \times 256 + 257 \times 128 + 128 \times 2 = 161408$
- D. $500 \times 256 \times 0.25 + 256 \times 128 \times 0.25 + 128 \times 2 = 40448$

Answer: D

NEW QUESTION 5

You developed an ML model with AI Platform, and you want to move it to production. You serve a few thousand queries per second and are experiencing latency issues. Incoming requests are served by a load balancer that distributes them across multiple Kubeflow CPU-only pods running on Google Kubernetes Engine (GKE). Your goal is to improve the serving latency without changing the underlying infrastructure. What should you do?

- A. Significantly increase the max_batch_size TensorFlow Serving parameter
- B. Switch to the tensorflow-model-server-universal version of TensorFlow Serving
- C. Significantly increase the max_enqueued_batches TensorFlow Serving parameter
- D. Recompile TensorFlow Serving using the source to support CPU-specific optimizations Instruct GKE to choose an appropriate baseline minimum CPU platform for serving nodes

Answer: A

NEW QUESTION 6

You are building a real-time prediction engine that streams files which may contain Personally Identifiable Information (PII) to Google Cloud. You want to use the Cloud Data Loss Prevention (DLP) API to scan the files. How should you ensure that the PII is not accessible by unauthorized individuals?

- A. Stream all files to Google CloudT and then write the data to BigQuery Periodically conduct a bulk scan of the table using the DLP API.
- B. Stream all files to Google Cloud, and write batches of the data to BigQuery While the data is being written to BigQuery conduct a bulk scan of the data using the DLP API.
- C. Create two buckets of data Sensitive and Non-sensitive Write all data to the Non-sensitive bucket Periodically conduct a bulk scan of that bucket using the DLP API, and move the sensitive data to the Sensitive bucket
- D. Create three buckets of data: Quarantine, Sensitive, and Non-sensitive Write all data to the Quarantine bucket.
- E. Periodically conduct a bulk scan of that bucket using the DLP API, and move the data to either the Sensitive or Non-Sensitive bucket

Answer: A

NEW QUESTION 7

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?

- 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

Answer: B

NEW QUESTION 8

You work with a data engineering team that has developed a pipeline to clean your dataset and save it in a Cloud Storage bucket. You have created an ML model and want to use the data to refresh your model as soon as new data is available. As part of your CI/CD workflow, you want to automatically run a Kubeflow Pipelines training job on Google Kubernetes Engine (GKE). How should you architect this workflow?

- A. Configure your pipeline with Dataflow, which saves the files in Cloud Storage After the file is saved, start the training job on a GKE cluster
- B. Use App Engine to create a lightweight python client that continuously polls Cloud Storage for new files As soon as a file arrives, initiate the training job
- C. Configure a Cloud Storage trigger to send a message to a Pub/Sub topic when a new file is available in a storage bucket
- D. Use a Pub/Sub-triggered Cloud Function to start the training job on a GKE cluster
- E. Use Cloud Scheduler to schedule jobs at a regular interval
- F. For the first step of the job
- G. check the timestamp of objects in your Cloud Storage bucket If there are no new files since the last run, abort the job.

Answer: A

NEW QUESTION 9

You have a functioning end-to-end ML pipeline that involves tuning the hyperparameters of your ML model using AI Platform, and then using the best-tuned parameters for training. Hypertuning is taking longer than expected and is delaying the downstream processes. You want to speed up the tuning job without significantly compromising its effectiveness. Which actions should you take?

Choose 2 answers

- A. Decrease the number of parallel trials
- B. Decrease the range of floating-point values
- C. Set the early stopping parameter to TRUE
- D. Change the search algorithm from Bayesian search to random search.
- E. Decrease the maximum number of trials during subsequent training phases.

Answer: DE

NEW QUESTION 10

You are designing an ML recommendation model for shoppers on your company's ecommerce website. You will use Recommendations AI to build, test, and deploy your system. How should you develop recommendations that increase revenue while following best practices?

- A. Use the "Other Products You May Like" recommendation type to increase the click-through rate
- B. Use the "Frequently Bought Together" recommendation type to increase the shopping cart size for each order.
- C. Import your user events and then your product catalog to make sure you have the highest quality event stream
- D. Because it will take time to collect and record product data, use placeholder values for the product catalog to test the viability of the model.

Answer: C

NEW QUESTION 10

You work for a social media company. You need to detect whether posted images contain cars. Each training example is a member of exactly one class. You have trained an object detection neural network and deployed the model version to AI Platform Prediction for evaluation. Before deployment, you created an evaluation job and attached it to the AI Platform Prediction model version. You notice that the precision is lower than your business requirements allow. How should you adjust the model's final layer softmax threshold to increase precision?

- A. Increase the recall
- B. Decrease the recall.
- C. Increase the number of false positives
- D. Decrease the number of false negatives

Answer: D

NEW QUESTION 15

You are training a TensorFlow model on a structured data set with 100 billion records stored in several CSV files. You need to improve the input/output execution performance. What should you do?

- 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)

Answer: B

NEW QUESTION 17

You work for an advertising company and want to understand the effectiveness of your company's latest advertising campaign. You have streamed 500 MB of campaign data into BigQuery. You want to query the table, and then manipulate the results of that query with a pandas dataframe in an AI Platform notebook. What should you do?

- A. Use AI Platform Notebooks' BigQuery cell magic to query the data, and ingest the results as a pandas dataframe
- B. Export your table as a CSV file from BigQuery to Google Drive, and use the Google Drive API to ingest the file into your notebook instance
- C. Download your table from BigQuery as a local CSV file, and upload it to your AI Platform notebook instance Use panda
- D. read_csv to ingest the file as a pandas dataframe
- E. From a bash cell in your AI Platform notebook, use the bq extract command to export the table as a CSV file to Cloud Storage, and then use gsutil cp to copy the data into the notebook Use panda
- F. read_csv to ingest the file as a pandas dataframe

Answer: B

NEW QUESTION 20

You work for a global footwear retailer and need to predict when an item will be out of stock based on historical inventory data. Customer behavior is highly dynamic since footwear demand is influenced by many different factors. You want to serve models that are trained on all available data, but track your performance on specific subsets of data before pushing to production. What is the most streamlined and reliable way to perform this validation?

- A. Use the TFX ModelValidator tools to specify performance metrics for production readiness
- B. Use k-fold cross-validation as a validation strategy to ensure that your model is ready for production.
- C. Use the last relevant week of data as a validation set to ensure that your model is performing accurately on current data
- D. Use the entire dataset and treat the area under the receiver operating characteristics curve (AUC ROC) as the main metric.

Answer: A

NEW QUESTION 24

You work for a large technology company that wants to modernize their contact center. You have been asked to develop a solution to classify incoming calls by product so that requests can be more quickly routed to the correct support team. You have already transcribed the calls using the Speech-to-Text API. You want to minimize data preprocessing and development time. How should you build the model?

- A. Use the AI Platform Training built-in algorithms to create a custom model
- B. Use AutoML Natural Language to extract custom entities for classification
- C. Use the Cloud Natural Language API to extract custom entities for classification
- D. Build a custom model to identify the product keywords from the transcribed calls, and then run the keywords through a classification algorithm

Answer: A

NEW QUESTION 25

You need to build classification workflows over several structured datasets currently stored in BigQuery. Because you will be performing the classification several times, you want to complete the following steps without writing code: exploratory data analysis, feature

selection, model building, training, and hyperparameter tuning and serving. What should you do?

- A. Configure AutoML Tables to perform the classification task
- B. Run a BigQuery ML task to perform logistic regression for the classification
- C. Use AI Platform Notebooks to run the classification model with pandas library
- D. Use AI Platform to run the classification model job configured for hyperparameter tuning

Answer: C

NEW QUESTION 28

You are building a linear model with over 100 input features, all with values between -1 and 1. You suspect that many features are non-informative. You want to remove the non-informative features from your model while keeping the informative ones in their original form. Which technique should you use?

- A. Use Principal Component Analysis to eliminate the least informative features.
- B. Use L1 regularization to reduce the coefficients of uninformative features to 0.
- C. After building your model, use Shapley values to determine which features are the most informative.
- D. Use an iterative dropout technique to identify which features do not degrade the model when removed.

Answer: C

NEW QUESTION 29

You are an ML engineer at a global shoe store. You manage the ML models for the company's website. You are asked to build a model that will recommend new products to the user based on their purchase behavior and similarity with other users. What should you do?

- A. Build a classification model
- B. Build a knowledge-based filtering model
- C. Build a collaborative-based filtering model
- D. Build a regression model using the features as predictors

Answer: C

NEW QUESTION 33

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?

- 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
- D. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources
- E. 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.

Answer: B

NEW QUESTION 36

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

Professional-Machine-Learning-Engineer Practice Exam Features:

- * Professional-Machine-Learning-Engineer Questions and Answers Updated Frequently
- * Professional-Machine-Learning-Engineer Practice Questions Verified by Expert Senior Certified Staff
- * Professional-Machine-Learning-Engineer Most Realistic Questions that Guarantee you a Pass on Your First Try
- * Professional-Machine-Learning-Engineer Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The Professional-Machine-Learning-Engineer Practice Test Here](#)