Vendor: Google

Exam Code: Professional-Machine-Learning-Engineer

Exam Name: Google Professional Machine Learning Engineer

Version: DEMO

QUESTION 1
You need to train a natural language model to perform text classification on product descriptions that contain millions of examples and 100,000 unique words. You want to preprocess the words individually so that they can be fed into a recurrent neural network. What should you do?

A. Create a hot-encoding of words, and feed the encodings into your model.
B. Identify word embeddings from a pre-trained model, and use the embeddings in your model.
C. Sort the words by frequency of occurrence, and use the frequencies as the encodings in your model.
D. Assign a numerical value to each word from 1 to 100,000 and feed the values as inputs in your model.

Answer: B
Explanation:
https://developers.google.com/machine-learning/guides/text-classification/ it is a Word Embedding case

QUESTION 2
You are using transfer learning to train an image classifier based on a pre-trained EfficientNet model. Your training dataset has 20,000 images. You plan to retrain the model once per day. You need to minimize the cost of infrastructure. What platform components and configuration environment should you use?

A. A Deep Learning VM with 4 V100 GPUs and local storage.
B. A Deep Learning VM with 4 V100 GPUs and Cloud Storage.
C. A Google Kubernetes Engine cluster with a V100 GPU Node Pool and an NFS Server
D. An AI Platform Training job using a custom scale tier with 4 V100 GPUs and Cloud Storage

Answer: D
Explanation:
The pre-trained EfficientNet model can be easily loaded from Cloud Storage, which eliminates the need for local storage or an NFS server. Using AI Platform Training allows for the automatic scaling of resources based on the size of the dataset, which can save costs compared to using a fixed-size VM or node pool. Additionally, the ability to use custom scale tiers allows for fine-tuning of resource allocation to match the specific needs of the training job.

QUESTION 3
While conducting an exploratory analysis of a dataset, you discover that categorical feature A has substantial predictive power, but it is sometimes missing. What should you do?

A. Drop feature A if more than 15% of values are missing. Otherwise, use feature A as-is.
B. Compute the mode of feature A and then use it to replace the missing values in feature A.
C. Replace the missing values with the values of the feature with the highest Pearson correlation with feature A.
D. Add an additional class to categorical feature A for missing values. Create a new binary feature that indicates whether feature A is missing.

Answer: D

QUESTION 4

You work for a large retailer and have been asked to segment your customers by their purchasing habits. The purchase history of all customers has been uploaded to BigQuery. You suspect that there may be several distinct customer segments, however you are unsure of how many, and you don't yet understand the commonalities in their behavior. You want to find the most efficient solution. What should you do?

A. Create a k-means clustering model using BigQuery ML. Allow BigQuery to automatically optimize the number of clusters.
B. Create a new dataset in Dataprep that references your BigQuery table. Use Dataprep to identify similarities within each column.
C. Use the Data Labeling Service to label each customer record in BigQuery. Train a model on your labeled data using AutoML Tables. Review the evaluation metrics to understand whether there is an underlying pattern in the data.
D. Get a list of the customer segments from your company's Marketing team. Use the Data Labeling Service to label each customer record in BigQuery according to the list. Analyze the distribution of labels in your dataset using Data Studio.

Answer: A
Explanation:
when to use k-means : Your data may contain natural groupings or clusters of data. You may want to identify these groupings descriptively in order to make data-driven decisions. For example, a retailer may want to identify natural groupings of customers who have similar purchasing habits or locations. This process is known as customer segmentation.
https://cloud.google.com/bigquery/docs/kmeans-tutorial

QUESTION 5
You recently designed and built a custom neural network that uses critical dependencies specific to your organization's framework. You need to train the model using a managed training service on Google Cloud. However, the ML framework and related dependencies are not supported by AI Platform Training. Also, both your model and your data are too large to fit in memory on a single machine. Your ML framework of choice uses the scheduler, workers, and servers distribution structure. What should you do?

A. Use a built-in model available on AI Platform Training.
B. Build your custom container to run jobs on AI Platform Training.
C. Build your custom containers to run distributed training jobs on AI Platform Training.
D. Reconfigure your code to a ML framework with dependencies that are supported by AI Platform Training.

Answer: C
Explanation:
By running your machine learning (ML) training job in a custom container, you can use ML frameworks, non-ML dependencies, libraries, and binaries that are not otherwise supported on Vertex AI.
Model and your data are too large to fit in memory on a single machine hence distributed training jobs.
https://cloud.google.com/vertex-ai/docs/training/containers-overview

QUESTION 6
While monitoring your model training's GPU utilization, you discover that you have a native synchronous implementation. The training data is split into multiple files. You want to reduce the execution time of your input pipeline. What should you do?

A. Increase the CPU load
B. Add caching to the pipeline
C. Increase the network bandwidth
D. Add parallel interleave to the pipeline

Answer: D
Explanation:
https://www.tensorflow.org/guide/data_performance

QUESTION 7
Your data science team is training a PyTorch model for image classification based on a pre-trained RestNet model. You need to perform hyperparameter tuning to optimize for several parameters. What should you do?

A. Convert the model to a Keras model, and run a Keras Tuner job.
B. Run a hyperparameter tuning job on AI Platform using custom containers.
C. Create a Kuberflow Pipelines instance, and run a hyperparameter tuning job on Katib.
D. Convert the model to a TensorFlow model, and run a hyperparameter tuning job on AI Platform.

Answer: B
Explanation:
Vertex AI supports custom models hyperparameter tuning.

QUESTION 8
You have a large corpus of written support cases that can be classified into 3 separate categories: Technical Support, Billing Support, or Other Issues. You need to quickly build, test, and deploy a service that will automatically classify future written requests into one of the categories. How should you configure the pipeline?

A. Use the Cloud Natural Language API to obtain metadata to classify the incoming cases.
B. Use AutoML Natural Language to build and test a classifier. Deploy the model as a REST API.
C. Use BigQuery ML to build and test a logistic regression model to classify incoming requests. Use BigQuery ML to perform inference.
D. Create a TensorFlow model using Google's BERT pre-trained model. Build and test a classifier, and deploy the model using Vertex AI.

Answer: B
Explanation:
AutoML is easier and faster and "you need to quickly build, test, and deploy". Also the REST API part fits our use case.

QUESTION 9
You need to quickly build and train a model to predict the sentiment of customer reviews with custom categories without writing code. You do not have enough data to train a model from scratch. The resulting model should have high predictive performance. Which service should you use?

A. AutoML Natural Language
B. Cloud Natural Language API

C. AI Hub pre-made Jupyter Notebooks
D. AI Platform Training built-in algorithms

Answer: A
Explanation:
https://cloud.google.com/natural-language/automl/docs/beginners-guide
The Natural Language API discovers syntax, entities, and sentiment in text, and classifies text into a predefined set of categories.

QUESTION 10
You need to build an ML model for a social media application to predict whether a user's submitted profile photo meets the requirements. The application will inform the user if the picture meets the requirements. How should you build a model to ensure that the application does not falsely accept a non-compliant picture?

A. Use AutoML to optimize the model's recall in order to minimize false negatives.
B. Use AutoML to optimize the model's F1 score in order to balance the accuracy of false positives and false negatives.
C. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that meet the profile photo requirements.
D. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that do not meet the profile photo requirements.

Answer: B
Explanation:
"Use AutoML to optimize the model's F1 score in order to balance the accuracy of false positives and false negatives", is the best approach to build an ML model that can predict whether a user's submitted profile photo meets the requirements while ensuring that the application does not falsely accept a non-compliant picture.

QUESTION 11
You lead a data science team at a large international corporation. Most of the models your team trains are large-scale models using high-level TensorFlow APIs on AI Platform with GPUs. Your team usually takes a few weeks or months to iterate on a new version of a model. You were recently asked to review your team's spending. How should you reduce your Google Cloud compute costs without impacting the model's performance?

A. Use AI Platform to run distributed training jobs with checkpoints.
B. Use AI Platform to run distributed training jobs without checkpoints.
C. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs with checkpoints.
D. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs without checkpoints.

Answer: C
Explanation:
https://cloud.google.com/blog/products/ai-machine-learning/reduce-the-costs-of-ml-workflows-with-preemptible-vms-and-gpus?hl=en

QUESTION 12
You need to train a regression model based on a dataset containing 50,000 records that is stored

in BigQuery. The data includes a total of 20 categorical and numerical features with a target variable that can include negative values. You need to minimize effort and training time while maximizing model performance. What approach should you take to train this regression model?

A. Create a custom TensorFlow DNN model
B. Use BQML XGBoost regression to train the model.
C. Use AutoML Tables to train the model without early stopping.
D. Use AutoML Tables to train the model with RMSLE as the optimization objective.

Answer: B

QUESTION 13
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 (PCA) 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: B
Explanation:
L1 regularization it's good for feature selection
https://www.quora.com/How-does-the-L1-regularization-method-help-in-feature-selection
https://developers.google.com/machine-learning/crash-course/regularization-for-sparsity/l1-regularization

QUESTION 14
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 then 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: C
Explanation:
https://www.tensorflow.org/tfx/guide/evaluator

QUESTION 15
You have deployed a model on Vertex AI for real-time inference. During an online prediction request, you get an "Out of Memory" error. What should you do?

A. Use batch prediction mode instead of online mode.
B. Send the request again with a smaller batch of instances.
C. Use base64 to encode your data before using it for prediction.
D. Apply for a quota increase for the number of prediction requests.

Answer: B
Explanation:
429 - Out of Memory
https://cloud.google.com/ai-platform/training/docs/troubleshooting

QUESTION 16
You work at a subscription-based company. You have trained an ensemble of trees and neural networks to predict customer churn, which is the likelihood that customers will not renew their yearly subscription. The average prediction is a 15% churn rate, but for a particular customer the model predicts that they are 70% likely to churn. The customer has a product usage history of 30%, is located in New York City, and became a customer in 1997. You need to explain the difference between the actual prediction, a 70% churn rate, and the average prediction. You want to use Vertex Explainable AI. What should you do?

A. Train local surrogate models to explain individual predictions.
B. Configure sampled Shapley explanations on Vertex Explainable AI.
C. Configure integrated gradients explanations on Vertex Explainable AI.
D. Measure the effect of each feature as the weight of the feature multiplied by the feature value.

Answer: B
Explanation:
Assigns credit for the outcome to each feature, and considers different permutations of the features. This method provides a sampling approximation of exact Shapley values.
shampled shapely recommended Model Type: Non-differentiable models, such as ensembles of trees and neural networks.
https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overview

QUESTION 17
You are working on a classification problem with time series data. After conducting just a few experiments using random cross-validation, you achieved an Area Under the Receiver Operating Characteristic Curve (AUC ROC) value of 99% on the training data. 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?

A. Address the model overfitting by using a less complex algorithm and use k-fold cross-validation.
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.

Answer: B
Explanation:
In your case, you are getting a very high AUC ROC value on the training data. This is a sign that

the model is overfitting. To address this, you should apply nested cross-validation during model training. This will help to prevent data leakage and ensure that the model is not overfitting to the training data.

QUESTION 18
You need to execute a batch prediction on 100 million records in a BigQuery table with a custom TensorFlow DNN regressor model, and then store the predicted results in a BigQuery table. You want to minimize the effort required to build this inference pipeline. What should you do?

A. Import the TensorFlow model with BigQuery ML, and run the ml.predict function.
B. Use the TensorFlow BigQuery reader to load the data, and use the BigQuery API to write the results to BigQuery.
C. Create a Dataflow pipeline to convert the data in BigQuery to TFRecords. Run a batch inference on Vertex AI Prediction, and write the results to BigQuery.
D. Load the TensorFlow SavedModel in a Dataflow pipeline. Use the BigQuery I/O connector with a custom function to perform the inference within the pipeline, and write the results to BigQuery.

Answer: A
Explanation:
https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-inference-overview
Predict the label, either a numerical value for regression tasks or a categorical value for classification tasks on DNN regresion.

QUESTION 19
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.

Answer: B
Explanation:
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.

Thank You for Trying Our Product

Braindump2go Certification Exam
Features:

★ More than **99,900** Satisfied Customers Worldwide.

★ Average **99.9%** Success Rate.

★ **Free Update** to match latest and real exam scenarios.

★ **Instant Download** Access! No Setup required.

★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.

★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.

★ **100%** Guaranteed Success or **100%** Money Back Guarantee.

★ **Fast**, helpful support **24x7**.

View list of all certification exams: http://www.braindump2go.com/all-products.html

10% Discount Coupon Code: ASTR14