Google Cloud

Feature Engineering

Quiz Question Answers

Module 1: Introduction to Vertex AI Feature Store

Question 1
What is one definition of a feature in machine learning?

A: A value that you receive from a model as an output
Feedback: This answer is incorrect, please review the module again.

B: A method of feature store
Feedback: This answer is incorrect, please review the module again.

*C: A value that is passed as input to a model
Feedback: This answer is correct.

D: A place to store any data
Feedback: This answer is incorrect, please review the module again.

Question 2
Vertex AI Feature Store provides a centralized repository for organizing, storing, and serving ML features. Using a central featurestore, enables an organization to efficiently share, discover, and re-use ML features at scale, which can increase the velocity of developing and deploying new ML applications. What are the key challenges that Vertex AI Feature Store solves?

A: Mitigate data storage silos, which occurs when you might have built and managed separate solutions for storage and the consumption of feature values.
Feedback: This answer is incorrect, please review the module again.

B: Detect drift, as a result of significant changes to your feature data distribution over time.
Feedback: This answer is partially correct, please review the module again.

C: Mitigate training-serving skew, which occurs when the feature data distribution that you use in production differs from the feature data distribution that was used to train your model.
Feedback: This answer is partially correct, please review the module again.

*D: All of the options are correct.

Feedback: This answer is correct.

Question 3
Where are the features registered?

*A: Feature registry
Feedback: This answer is correct.

B: Online Store
Feedback: This answer is incorrect, please review the module again.

C: Offline Store
Feedback: This answer is incorrect, please review the module again.

D: Feature Monitoring
Feedback: This answer is incorrect, please review the module again.

Question 4
Which of the following is an instance of an entity type?

A: Feature
Feedback: This answer is incorrect, please review the module again.

B: Online Store
Feedback: This answer is incorrect, please review the module again.

*C: Entity
Feedback: This answer is correct.

D: Featurestore
Feedback: This answer is incorrect, please review the module again.

Question 5
What are the two methods feature store offers for serving features?

A: Online serving and Offline serving
Feedback: This answer is incorrect, please review the module again.

*B: Batch serving and Online serving
Feedback: This answer is correct.

C: Offline serving and Stream serving
Feedback: This answer is incorrect, please review the module again.

D: Batch serving and Stream serving
Feedback: This answer is incorrect, please review the module again.

Question 6
Which of the following is the process of importing feature values computed by your feature engineering jobs into a featurestore?

A: Feature store
Feedback: This answer is incorrect, please review the module again.

B: Feature Monitoring
Feedback: This answer is incorrect, please review the module again.

C: Feature serving
Feedback: This answer is incorrect, please review the module again.

*D: Feature ingestion
Feedback: This answer is correct.

Module 2: Raw Data to Features

Question 1
In what form can raw data be used inside ML models?

*A: After turning your raw data into a useful feature vectors
Feedback: This answer is correct.

B: After turning your raw data into a useful feature matrix
Feedback: This answer is incorrect, please review the module again.

C: After turning your raw data into multidimensional vectors
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 2
A good feature has which of the following characteristics?

A: It should be related to the objective.
Feedback: This answer is partially correct, please review the module again.

B: It should be known at prediction time.
Feedback: This answer is partially correct, please review the module again.

C: It should be numeric with meaningful magnitude.
Feedback: This answer is partially correct, please review the module again.

*D: All of the options are correct.
Feedback: This answer is correct.

Question 3
Which of the following are the requirements to build an effective machine learning model?

A: It should scale to a large dataset.
Feedback: This answer is partially correct, please review the module again.

B: It should find good features.
Feedback: This answer is partially correct, please review the module again.

C: It should be able to preprocess with Vertex AI Platform.
Feedback: This answer is partially correct, please review the module again.

*D: All of the options are correct.
Feedback: This answer is correct

Question 4
Which of the following statements is true about preprocessing?

*A: Preprocessing within the context of Cloud ML allows you to do it at scale.
Feedback: This answer is correct.

B: Preprocessing without the context of Cloud ML allows you to do it at scale.
Feedback: This answer is incorrect, please review the module again.

C: Both options are correct.
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 5
Which of the following statements is true?

A: Same problems in the same domain may need different features.
Feedback: This answer is incorrect, please review the module again.

*B: Different problems in the same domain may need different features.
Feedback: This answer is correct

C: Different problems in different domains may need the same features.
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback:This answer is incorrect, please review the module again.

Module 3: Feature Engineering

Question 1
True or False: Feature Engineering is often one of the most valuable tasks a data scientist can do to improve model performance, for three main reasons:

1. You can isolate and highlight key information, which helps your algorithms "focus" on what's important.
2. You can bring in your own domain expertise.
3. Once you understand the "vocabulary" of feature engineering, you can bring in other people's domain expertise.

*A: True
Feedback: This answer is correct.

B: False
Feedback: This answer is incorrect, please review the module again.

Question 2
What is one-hot encoding?

*A: One-hot encoding is a process by which categorical variables are converted into a form that could be provided to neural networks to do a better job in prediction.
Feedback: This answer is correct..

B: One-hot encoding is a process by which numeric variables are converted into a form that could be provided to neural networks to do a better job in prediction.
Feedback: This answer is incorrect, please review the module again.

C: One-hot encoding is a process by which numeric variables are converted into a categorical form that could be provided to neural networks to do a better job in prediction.
Feedback: This answer is incorrect, please review the module again.

D: One-hot encoding is a process by which only the hottest numeric variable is retained for use by the neural network.
Feedback: This answer is incorrect, please review the module again.

Question 3
What do you use the tf.feature_column.bucketized_column function for?

A: To compute the hash buckets needed to one-hot encode categorical values
Feedback: This answer is incorrect, please review the module again.

B: To count the number of unique buckets the input values falls into
Feedback: This answer is incorrect, please review the module again.

*C: To discretize floating point values into a smaller number of categorical bins
Feedback: This answer is correct.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 4
What is a feature cross?

A: A feature cross is a synthetic feature formed by adding (crossing) two or more features. Crossing combinations of features can provide predictive abilities beyond what those features can provide individually.
Feedback: This answer is incorrect, please review the module again.

*B: A feature cross is a synthetic feature formed by multiplying (crossing) two or more features. Crossing combinations of features can provide predictive abilities beyond what those features can provide individually.

Feedback: This answer is correct.

C: A feature cross is a synthetic feature formed by dividing (crossing) two or more features. Crossing combinations of features can provide predictive abilities beyond what those features can provide individually.
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 5:
Which of the following statements are true regarding the ML.EVALUATE function?

A: The ML.EVALUATE function can be used with linear regression, logistic regression, k-means, matrix factorization, and ARIMA-based time series models.
Feedback: This answer is partially correct, please review the module again.

B: The ML.EVALUATE function evaluates the predicted values against the actual data.
Feedback: This answer is partially correct, please review the module again.

C: You can use the ML.EVALUATE function to evaluate model metrics.
Feedback: This answer is partially correct, please review the module again.

*D: All of the options are correct.
Feedback: This answer is correct.

Question 6:
What is the significance of ML.FEATURE_CROSS?

*A: ML.FEATURE_CROSS generates a STRUCT feature with all combinations of crossed categorical features except for 1-degree items.
Feedback: This answer is correct.

B: ML.FEATURE_CROSS generates a STRUCT feature with few combinations of crossed categorical features except for 1-degree items.
Feedback: This answer is incorrect, please review the module again.

C: ML.FEATURE_CROSS generates a STRUCT feature with all combinations of crossed categorical features including 1-degree items.
Feedback: This answer is incorrect, please review the module again.

D: ML.FEATURE_CROSS generates a STRUCT feature with few combinations of crossed categorical features except for 1-degree items.
Feedback: This answer is incorrect, please review the module again.

Question 7:
Which of the following statements are true regarding the ML.BUCKETIZE function?

A: ML.BUCKETIZE is a pre-processing function that creates buckets by returning a STRING as the bucket name after numerical_expression is split into buckets by array_split_points..
Feedback: This answer is partially correct, please review the module again.

B: It bucketizes a continuous numerical feature into a string feature with bucket names as the value.
Feedback: This answer is partially correct, please review the module again.

*C: Both options are correct.
Feedback: This answer is correct.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 8:
Which of the following is true about Feature Cross?

A: It is a process of combining features into a single feature.
Feedback: This answer is partially correct, please review the module again.

B: Feature Cross enables a model to learn separate weights for each combination of features.
Feedback: This answer is partially correct, please review the module again.

*C: Both options are correct.
Feedback: This answer is correct.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Module 4: Preprocessing and feature creation

Question 1

Which of these accurately describes the relationship between Apache Beam and Cloud Dataflow?

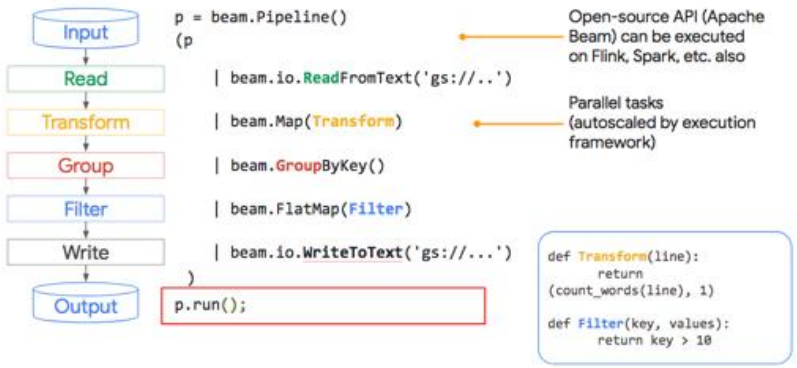
*A: Cloud Dataflow is the API for data pipeline building in java or python and Apache Beam is the implementation and execution framework.
Feedback: This answer is correct.

B: They are the same.
Feedback: This is incorrect, please review the module again.

C: Cloud Dataflow is the proprietary version of the Apache Beam API and the two are not compatible.
Feedback: This is incorrect, please review the module again.

Question 2

True or False: The Filter method can be carried out in parallel and autoscaled by the execution framework:



*A: True: Anything in Map or FlatMap can be parallelized by the Beam execution framework.
Feedback: This answer is correct.

B: False: Anything in Map or FlatMap can be parallelized by the Beam execution framework.
Feedback: This is incorrect, please review the module again.

Question 3
What is the purpose of a Cloud Dataflow connector?

```
.apply(TextIO.write().to("gs://..."));
```

*A: Connectors allow you to output the results of a pipeline to a specific data sink like Bigtable, Google Cloud Storage, flat file, BigQuery, and more.
Feedback: This answer is correct.

B: Connectors allow you to chain multiple data-processing steps together automatically so they process in parallel.
Feedback: This is incorrect, please review the module again.

C: Connectors allow you to authenticate your pipeline as specific users who may have greater access to datasets.
Feedback: This is incorrect, please review the module again.

Question 4
To run a pipeline you need something called a _____.

*A: runner
Feedback: This answer is correct.

B: executor
Feedback: This is incorrect, please review the module again.

C: pipeline
Feedback: This is incorrect, please review the module again.

D: Apache Beam
Feedback: This is incorrect, please review the module again.

Question 5
Your development team is about to execute this code block. What is your team about to do?

```
mvn compile -e exec:java \
      -Dexec.mainClass=$MAIN \
      -Dexec.args="--project=$PROJECT \
      --stagingLocation=gs://$BUCKET/staging/ \
      --tempLocation=gs://$BUCKET/staging/ \
      --runner=DataflowRunner"
```

*A: We are compiling our Cloud Dataflow pipeline written in Java and are submitting it to the cloud for execution.
Notice that we are calling mvn compile and passing in --runner=DataflowRunner.
Feedback: This answer is correct.

B: We are compiling our Cloud Dataflow pipeline written in Python and are loading the outputs of the executed pipeline inside of Google Cloud Storage (gs://)
Feedback: This is incorrect, please review the module again.

C: We are preparing a staging area in Google Cloud Storage for the output of our Cloud Dataflow pipeline and will be submitting our BigQuery job with a later command.
Feedback: This is incorrect, please review the module again.

Question 6
True or False: A ParDo acts on all items at once (like a Map in MapReduce).

A: True
Feedback: This is incorrect, please review the module again.

*B: False. A ParDo acts on one item at a time (like a Map in MapReduce)
Feedback: This answer is correct.

Question 7
What is one key advantage of preprocessing your features using Apache Beam?

*A: The same code you use to preprocess features in training and evaluation can also be used in serving.
Feedback: This answer is correct.

B: Apache Beam transformations are written in Standard SQL which is scalable and easy to author.

Feedback: This is incorrect, please review the module again.

C: Apache Beam code is often harder to maintain and run at scale than BigQuery preprocessing pipelines.
Feedback: This is incorrect, please review the module again.

Module 5: Feature Crosses - TensorFlow Playground

Question 1
True or False: We can create many different kinds of feature crosses. For example:

- [A X B]: a feature cross formed by multiplying the values of two features.
- [A x B x C x D x E]: a feature cross formed by multiplying the values of five features.
- [A x A]: a feature cross formed by squaring a single feature.

*A: True
Feedback: This answer is correct.

B: False
Feedback: This answer is incorrect, please review the module again.

Question 2
True or False: In TensorFlow Playground, orange and blue are used throughout the visualization in slightly different ways, but in general orange shows negative values while blue shows positive values.

*A: True
Feedback: This answer is correct.

B: False
Feedback: This answer is incorrect, please review the module again.

Question 3
True or False: In TensorFlow Playground, the data points (represented by small circles) are initially colored orange or blue, which correspond to zero and negative one.

A: True
Feedback: This answer is incorrect, please review the module again.

*B: False

Feedback: This answer is correct. The answer is positive one to negative one.

Question 4

Fill in the blanks: In the ____ layers, the lines are colored by the _____ of the connections between neurons. Blue shows a _____ weight, which means the network is using that ____ of the neuron as given. An orange line shows that the network is assigning a _____ weight.

A: Hidden, weights, negative, output, positive
Feedback: This answer is incorrect, please review the module again.

*B: Hidden, weights, positive, output, negative
Feedback: This answer is correct.

C: Weights, hidden, negative, output, positive
Feedback: This answer is incorrect, please review the module again.

D: Output, weights, negative, hidden, positive
Feedback: This answer is incorrect, please review the module again.

Question 5

True or False: In TensorFlow Playground, in the output layer, the dots are colored orange or blue depending on their original values. The background color shows what the network is predicting for a particular area. The intensity of the color shows how confident that prediction is.

*A: True
Feedback: This answer is correct..

B: False
Feedback: This answer is incorrect, please review the module again.

Question 6

Why might you create an embedding of a feature cross?

A. To create a lower-dimensional representation of the input space
Feedback: This is one of the correct answers.

B. To identify similar sets of inputs for clustering
Feedback: This is one of the correct answers.

C. To reuse weights learned in one problem in another problem
Feedback: This is one of the correct answers.

*D. All of the options are correct.
This answer is correct.

Module 6: Introduction to TensorFlow Transform

Question 1
What does tf.Transform do during the training and serving phase?

*A: Provides a TensorFlow graph for preprocessing
Feedback: This answer is correct.

B: Provides computation over the entire dataset, including on both internal and external data sources
Feedback: This answer is incorrect, please review the module again.

C: Provides a transformation polynomial to train the data
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 2
What is Tensorflow Transform a hybrid of?

*A: Apache and TensorFlow
Feedback: This answer is correct.

B: Dataflow and Tensorflow
Feedback: This answer is incorrect, please review the module again.

C: Both options are correct.
Feedback: This answer is incorrect, please review the module again.

D: None of the options are correct.
Feedback: This answer is incorrect, please review the module again.

Question 3

True or False: One of the goals of tf.Transform is to provide a TensorFlow graph for preprocessing that can be incorporated into the serving graph (and, optionally, the training graph).

*A: True
Feedback: This answer is correct.

B: False
Feedback: This answer is incorrect, please review the module again.

Question 4
Fill in the blank:
The _____ _____ is the most important concept of tf.Transform. The _____ _____ is a logical description of a transformation of the dataset. The _____ _____ accepts and returns a dictionary of tensors, where a tensor means Tensor or 2D SparseTensor.

*A: Preprocessing function
Feedback: This answer is correct.

B: Preprocessing variable
Feedback: This answer is incorrect, please review the module again.

C: Preprocessing method
Feedback: This answer is incorrect, please review the module again.