

Get to know your data: Improve data through Exploratory Data Analysis

Question 1

Which of the following are categories of data quality tools?

- A. Cleaning tools
- B. Monitoring tools
- C. Both 'Cleaning tools' and 'Monitoring tools'
- D. None of the options

Question 2

What are the features of low data quality?

- A. Unreliable info
- B. Incomplete data
- C. Duplicated data
- D. All of the options

Question 3

What are the objectives of exploratory data analysis?

- A. Check for missing data and other mistakes.
- B. Gain maximum insight into the data set and its underlying structure.
- C. Uncover a parsimonious model, one which explains the data with a minimum number of predictor variables.
- D. All of the options

Question 4

Exploratory Data Analysis is majorly performed using the following methods:

- A. Univariate
- B. Bivariate
- C. Both Univariate and Bivariate
- D. None of the options

Question 5

Which of the following is not a component of Exploratory Data Analysis?

- A. Accounting and Summarizing
- B. Anomaly Detection
- C. Statistical Analysis and Clustering
- D. Hyperparameter tuning

Machine Learning in Practice

Question 1

Which of the following machine learning models have labels, or in other words, the correct answers to whatever it is that we want to learn to predict?

- A. Unsupervised Model
- B. Supervised Model
- C. Reinforcement Model
- D. None of the options

Question 2

Which model would you use if your problem required a discrete number of values or classes?

- A. Regression Model
- B. Unsupervised Model
- C. Supervised Model
- D. Classification Model

Question 3

To predict the continuous value of our label, which of the following algorithms is used?

- A. Classification
- B. Regression
- C. Unsupervised
- D. None of the options

Question 4

What is the most essential metric a regression model uses?

- A. Mean squared error as their loss function
- B. Cross entropy
- C. Both 'Mean squared error as their loss function' & 'Cross entropy'
- D. None of the options

Question 5

Why is regularization important in logistic regression?

- A. Avoids overfitting
- B. Keeps training time down by regulating the time allowed
- C. Finds errors in the algorithm
- D. Encourages the use of large weights

Training AutoML Models Using Vertex AI

Question 1

What is the main benefit of using an automated Machine Learning workflow?

- A. It makes the model perform better.
- B. It makes the model run faster.
- C. It deploys the model into production.
- D. It reduces the time it takes to develop trained models and assess their performance.

Question 2

For a user who can use SQL, but has little Machine Learning experience and wants a 'Low-Code' solution, which Machine Learning framework should they use?

- A. Scikit-Learn
- B. BigQuery ML
- C. AutoML
- D. Python

Question 3

If a dataset is presented in a Comma Separated Values (CSV) file, which is the correct data type to choose in Vertex AI?

- A. Image
- B. Tabular**
- C. Text
- D. Video

Question 4

If the business case is to predict fraud detection, which is the correct Objective to choose in Vertex AI?

- A. Clustering
- B. Forecasting
- C. Regression/Classification**
- D. Segmentation

Question 5

Which of the following are stages of the Machine Learning workflow that can be managed with Vertex AI?

- A. Create a dataset and upload data.
- B. Train an ML model on your data.
- C. Deploy your trained model to an endpoint for serving predictions.
- D. All of the options.**

Question 6

What is the default setting in AutoML Tables for the data split in model evaluation?

- A. 70% Training, 20% Validation, 10% Testing
- B. 80% Training, 15% Validation, 5% Testing
- C. 80% Training, 5% Validation, 15% Testing
- D. 80% Training 10% Validation, 10% Testing**

Question 7

MAE, MAPE, RMSE, RMSLE and R2 are all available as test examples in the Evaluate section of Vertex AI and are common examples of what type of metric?

- A. Forecasting Regression Metrics
- B. Linear Regression Metrics**
- C. Clustering Regression Metrics
- D. Decision Trees Progression Metrics

Question 8

What does the Feature Importance attribution in Vertex AI display?

- A. How much each feature impacts the model, expressed as a percentage
- B. How much each feature impacts the model, expressed as a ratio
- C. How much each feature impacts the model, expressed as a ranked list
- D. How much each feature impacts the model, expressed as a decimal

Question 9

Which of the following metrics can be used to find a suitable balance between precision and recall in a model?

- A. PR AUC
- B. ROC AUC
- C. Log Loss
- D. F1 Score**

BigQuery Machine Learning: Develop ML Models Where Your Data Lives

Question 1

Which of the following are advantages of BigQuery ML when compared to Python based ML frameworks?

- A. BigQuery ML custom models can be created without the use of multiple tools
- B. BigQuery ML automates multiple steps in the ML workflow
- C. Moving and formatting large amounts of data takes longer with Python based models compared to model training in BigQuery
- D. All of the options**

Question 2

Which of these BigQuery supported classification models is most relevant for predicting binary results, such as True/False?

- A. XGBoost
- B. AutoML Tables
- C. Logistic Regression
- D. DNN Classifier (TensorFlow)

Question 3

For Classification or Regression problems with decision trees, which of the following models is most relevant?

- A. Wide and Deep NNs
- B. AutoML Tables
- C. Linear Regression
- D. XGBoost

Question 4

Where labels are not available, for example where customer segmentation is required, which of the following BigQuery supported models is useful?

- A. Time Series Forecasting
- B. Recommendation - Matrix Factorization
- C. K-Means Clustering
- D. Time Series Anomaly Detection

Question 5

What are the 3 key steps for creating a Recommendation System with BigQuery ML?

- A. Prepare training data in BigQuery, train a recommendation system with BigQuery ML, use the predicted recommendations in production
- B. Import training data to BigQuery, train a recommendation system with BigQuery ML, tune the hyperparameters
- C. Prepare training data in BigQuery, select a recommendation system from BigQuery ML, deploy and test the model
- D. Prepare training data in BigQuery, specify the model options in BigQuery ML, export the predictions to Google Analytics

Optimization

Question 1

For the formula used to model the relationship i.e. $y = mx + b$, what does 'm' stand for?

- A. It captures the amount of change we've observed in our label in response to a small change in our feature.
- B. It refers to a bias term which can be used for regression.
- C. It refers to a bias term which can be used for regression and it captures the amount of change we've observed in our label in response to a small change in our feature.
- D. None of the options are correct.

Question 2

What are the basic steps in an ML workflow (or process)?

- A. Collect data
- B. Check for anomalies, missing data and clean the data
- C. Perform statistical analysis and initial visualization
- D. All options are correct.

Question 3

Which of the following loss functions is used for classification problems?

- A. MSE
- B. Cross entropy
- C. Both MSE & Cross entropy
- D. None of the options are correct.

Question 4

Which of the following gradient descent methods is used to compute the entire dataset?

- A. Batch gradient descent
- B. Gradient descent
- C. Mini-batch gradient descent
- D. None of the options are correct.

Question 5

Which of the following are benefits of Performance metrics over loss functions?

- A. Performance metrics are easier to understand.
- B. Performance metrics are directly connected to business goals.
- C. Performance metrics are easier to understand and are directly connected to business goals.
- D. None of the options are correct.

Generalization and Sampling

Question 1

Which is the best way to assess the quality of a model?

- A. Observing how well a model performs against a new dataset that it hasn't seen before.
- B. Observing how well a model performs against an existing known dataset.
- C. None of the options are correct.
- D. Observing how well a model performs against a new dataset that it hasn't seen before and observing how well a model performs against an existing known dataset.

Question 2

How do you decide when to stop training a model?

- A. When your loss metrics start to increase
- B. When your loss metrics start to both increase and decrease
- C. When your loss metrics start to decrease
- D. None of the options are correct

Question 3

Which of the following actions can you perform on your model when it is trained and validated?

- A. You can write it multiple times against the dependent test dataset.
- B. You can write it once, and only once against the dependent test dataset.
- C. You can write it multiple times against the independent test dataset.
- D. You can write it once, and only once, against the independent test dataset.

Question 4

Which of the following allows you to create repeatable samples of your data?

- A. Use the last few digits of a hash function on the field that you're using to split or bucketize your data.
- B. None of the options are correct.
- C. Use the first few digits of a hash function on the field that you're using to split or bucketize your data.
- D. Use the first few digits or the last few digits of a hash function on the field that you're using to split or bucketize your data.

Question 5

Which of the following allows you to split the dataset based upon a field in your data?

- A. BUCKETIZE, an open-source hashing algorithm that is implemented in BigQuery SQL.
- B. None of the options are correct.
- C. ML_FEATURE FINGERPRINT, an open-source hashing algorithm that is implemented in BigQuery SQL.
- D. FARM_FINGERPRINT, an open-source hashing algorithm that is implemented in BigQuery SQL.