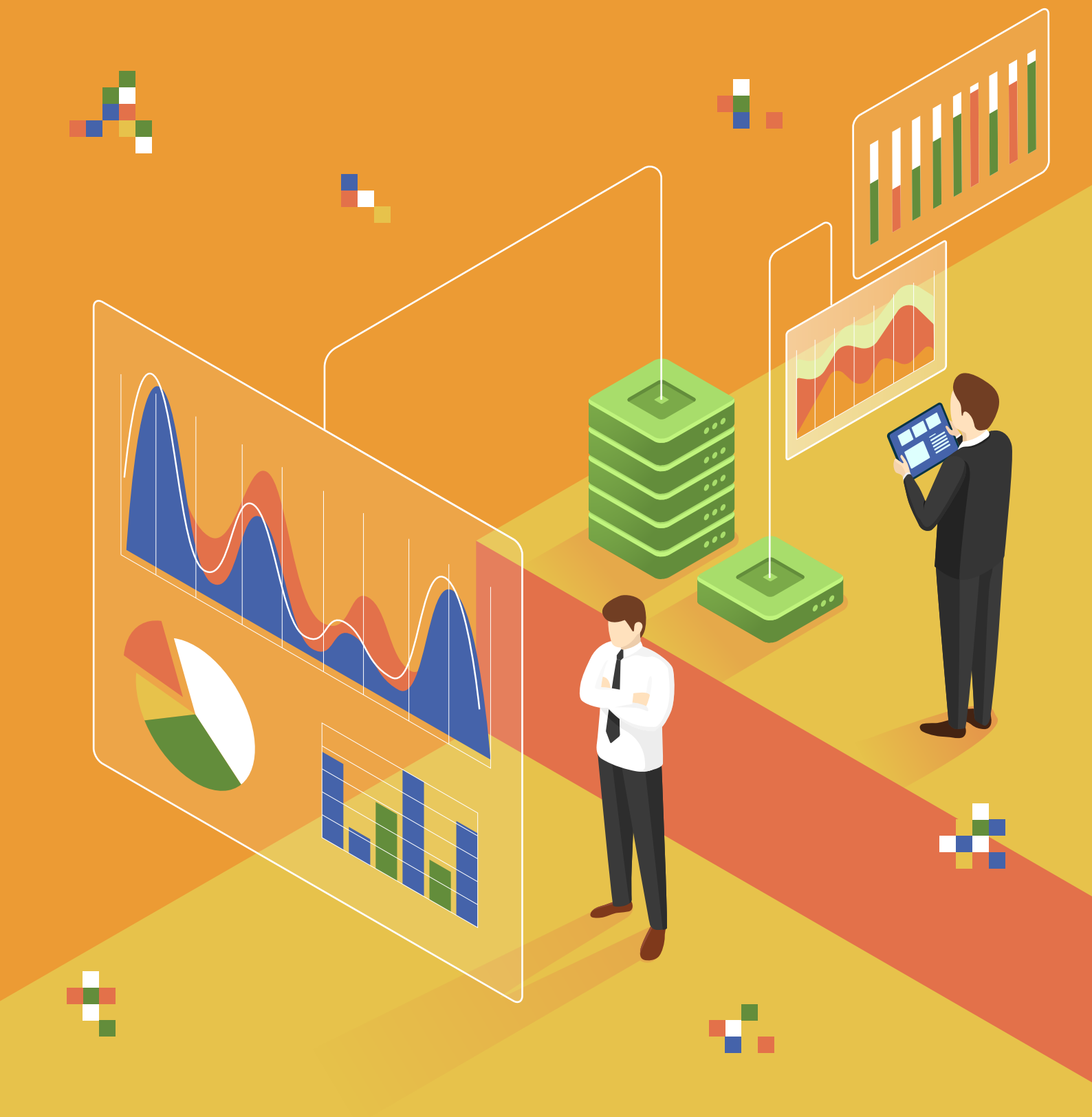
SuperDataScience

10 SAMPLE EXAM QUESTIONS
+ COMPLETE ANSWER KEY



MACHINE AND DEEP LEARNING BASICS — PART #2

1. **Your company is working on a project that helps in election campaign. You are tasked to build a ML model to predict the outcome after elections in several states. Your company has collected data from different states with majority of data coming from states which is under the XY party. After reviewing the results, the model is found to biased. What is the problem here?**

   A Data collected has missing features

   B Data collected has few observations and it is making a bias prediction

   C Data should be collected equally from all the states, not only from States under XY party

   D There is no problem with data and we need to tune hyper-parameters

2. **You are tasked to create a ML model for car sale company to predict the potential customers. You are given access to data which contains earning per individual which ranges from 1000 dollars to 2 million dollars, net worth, number of cars they own and number of family members. Initial analysis shows there are no outliers in net worth and earnings. But after using ML model, you model seems to perform poorly. How can you improve your model?**

   A Normalize the data of net worth and earnings

   B Standardize the data of net worth and earnings

   C Apply categorical encoding

   D Apply one-hot encoding

3. **You are working for an analytic company and you are presenting the outcome of your model using a confusion matrix. Your model has poor recall, which error might affect recall the most?**

   A Type I Error
   B Type II Error
   C Type III Error
   D Type IV Error

4. **You have been given access to student marks data and you want to predict the top performing students using AWS Linear Learner. On training you find that the model is overfitting the data, which metric be used to tune the model so as to avoid overfitting?**

   A Training metric
   B Overfitting metric
   C Underfitting metric
   D Validation metric

5. **You are creating your own image classification algorithm and you want to check the training metrics if anything goes wrong while training for debugging purpose using CloudWatch. What do you have to do to so that training metrics are available in CloudWatch?**

   A You have to make sure relevant metrics from stdout and stderr streams are included in the metric_definition while defining the training job
   B Sagemaker will automatically add the metrics from stdout and stderr streams to the metric_definition while defining the training job

C You can use cloud-trail to track the metrics

D You cannot log the metrics for a custom defined algorithm

6. **You are creating image classification model to classify 4 different objects like car, bike, pedestrians and truck. After training the model, you receive a Macro F1 Score of 0.8 and the F1 score for pedestrian class is 0.3. What does this mean?**

A It is due to type I error

B It might be due to the presence of more training data for pedestrians compared to that of other classes

C It might be due to the presence of less training data for pedestrians compared to that of other classes

D It is due to type II error

7. **You are creating a fraud-detection system for the bank. Bank does not want to lose money at any case to the fraudulent activities. This is their higher priority. Which type of error should be minimum for this case?**

A Type I Error

B Type II Error

C Type III Error

D Type IV Error

8. **You are creating ML model for a time-series prediction to replace the existing model in your company's application and would like to do offline validation before exposing to the live production traffic. Which method will help you with the validation?**

A Use canary method with historical data

B Use A/B test to expose the model to simulated traffic

C Use rolling deployment by exposing the model to simulated traffic

D Use back-testing with historical data

9. You are building a stock price prediction model using deep learning. What type of neural network would help in this type of problem?

A CNN

B LSTM

C MLP

D K-NN

10. A classifier that predicts if a patient has cancer or not is represented by the confusion matrix. Confusion matrix has the following values: True Positive = 20, False Positive = 80, False Negative = 40 and True Negative = 70. What is the recall of this classifier?

A 0.33

B 0.45

C 0.6

D 0.2

ANSWER KEY:

1. C	2. A	3. B	4. D	5. A
6. C	7. B	8. D	9. B	10. A