Vendor:IBM

Exam Code:C1000-059

Exam Name:IBM AI Enterprise Workflow V1 Data Science Specialist

Version: Demo

QUESTION 1

What are two methods used to detect outliers in structured data? (Choose two.)

A. multi-label classification

B. isolation forest

C. gradient descent

D. one class Support Vector Machine (SVM)

E. Word2Vec

Correct Answer: BD

Reference: https://www.researchgate.net/post/What-is-the-best-outliers-detection-algorithm-to-used-for- big-data

---

QUESTION 2

What are two hyperparameters used when building a k-means model? (Choose two.)

A. kernel

B. learning rate

C. number of iterations

D. number of clusters

E. number of neighbors

Correct Answer: CD

Reference: https://www.ritchieng.com/machine-learning-clustering-kmeans/

QUESTION 3

A new test to diagnose a disease is evaluated on 1152 people, and 106 people have the disease, and 1046 people do not have the disease. The test results are summarized below:

	Test predicts disease	Test predicts no disease
Have the disease	73	33
Do not have the disease	81	965

In this sample, how many cases are false positives and false negatives?

A. 33 false positives and 81 false negatives

B. 81 false positives and 73 false negatives

C. 73 false positives and 81 false negatives

D. 81 false positives and 33 false negatives

Correct Answer: A

QUESTION 4

What is the meaning of "deep" in deep learning?

A. To go deep into the loss function landscape.

B. The higher the number of machine learning algorithms that can be applied, the deeper is the learning.

C. A kind of deeper understanding achieved by any approach taken.

D. It indicates the many layers contributing to a model of the data.

Correct Answer: D

Reference: https://en.wikipedia.org/wiki/Deep_learning

QUESTION 5

Considering one ML application is deployed using Kubernetes, its output depends on the data which is constantly stored in the model, if needing to scale the system based on available CPUs, what feature should be enabled?

A. persistent storage

B. vertical pod autoscaling

C. horizontal pod autoscaling

D. node self-registration mode

Correct Answer: A

QUESTION 6

A data analyst creates a term-document matrix for the following sentence:

I saw a cat, a dog and another cat.

Assuming they used a binary vectorizer, what is the resulting weight for the word cat?

A. 0

B. 1

C. 3

D. 2

Correct Answer: B

QUESTION 7

The formula for recall is given by (True Positives) / (True Positives + False Negatives).

What is the recall for this example?

	predicted	
	negative	**positive**
negative	3	2
positive	4	1

actual (label for rows: negative, positive)

A. 0.2

B. 0.25

C. 0.5

D. 0.33

Correct Answer: B

Reference: https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced- classification/

QUESTION 8

Which of the following entity extraction techniques would be best for the extraction of telephone numbers from a text document?

A. complex pattern-based

B. regex

C. statistical

D. dictionary

Correct Answer: C

Reference: https://www.researchgate.net/ publication/318093829_Developing_an_innovative_entity_extraction_method_for_unstructured_data


QUESTION 9

Determine the number of bigrams and trigrams in the sentence. "Data is the new oil".

A. 3 bigrams, 3 trigrams

B. 4 bigrams, 4 trigrams

C. 3 bigrams, 4 trigrams

D. 4 bigrams, 3 trigrams

Correct Answer: A

QUESTION 10

A neural network is trained for a classification task. During training, you monitor the loss function for the train dataset and the validation dataset, along with the accuracy for the validation dataset. The goal is to get an accuracy of 95%.
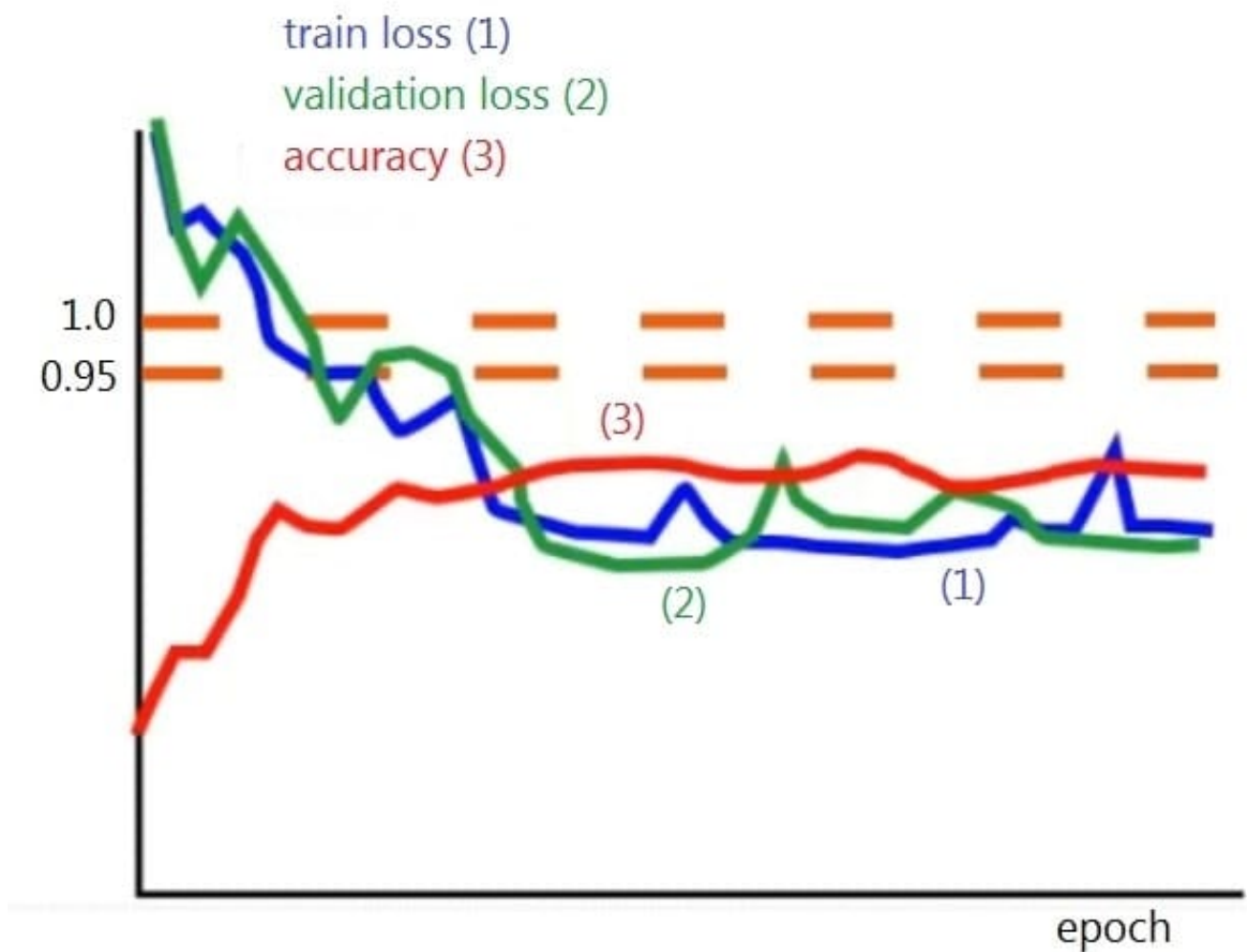
From the graph, what modification would be appropriate to improve the performance of the model?

A. increase the depth of the neural network

B. insert a dropout layer in the neural network architecture

C. increase the proportion of the train dataset by moving examples from the validation dataset to the train dataset

D. restart the training with a higher learning rate

Correct Answer: D

QUESTION 11

What is the goal of the backpropagation algorithm?

A. to randomize the trajectory of the neural network parameters during training

B. to smooth the gradient of the loss function in order to avoid getting trapped in small local minimas

C. to scale the gradient descent step in proportion to the gradient magnitude

D. to compute the gradient of the loss function with respect to the neural network parameters

Correct Answer: B

Reference: https://www.sciencedirect.com/topics/computer-science/backpropagation

---

QUESTION 12

Given the following sentence:

The dog jumps over a fence.

What would a vectorized version after common English stopword removal look like?

A. [\\'dog\\', \\'fence\\', \\'run\\']

B. [\\'fence\\', \\'jumps\\']

C. [\\'dog\\', \\'fence\\', \\'jumps\\']

D. [\\'a\\', \\'dog\\', \\'fence\\', \\'jumps\\', \\'over\\', \\'the\\']

Correct Answer: C

Reference: https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different- libraries-f20bac19929a