Jane Chinwuko Project: Batch DS2307 PFA file 3.
1) Among the following identify the one in which dimensionality reduction reduces.
a) Performance
b) statistics
c) Entropy
d) Collinearity
Answer is D
2) Which of the following machine learning algorithm is based upon the idea of bagging?
a) Decision Tree
b) Random Forest
c) Classfication
d) SVM
Answer is B
3) Choose a disadvantage of decision trees among the following.
a) Decision tree robust to outliers
b) Factor analysis
c) Decision Tree are prone to overfit
d) all of the above
Answer is C
<b>4)</b> What is the term known as on which the machine learning algorithms build a model based on sample data?
a) Data Training
b) Sample Data
c) Training data

Answer is C

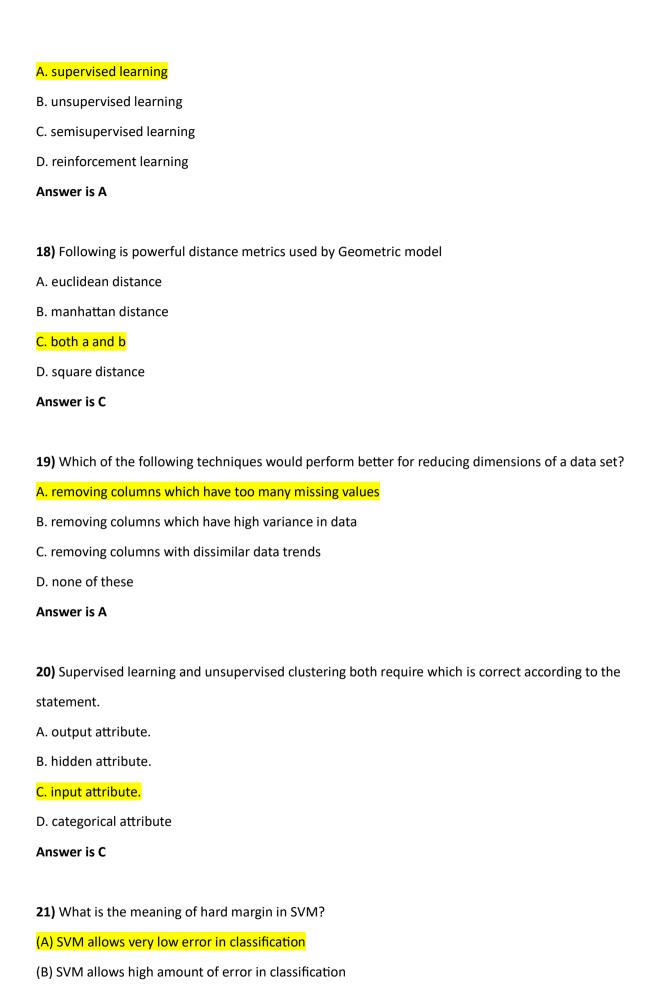
d) None of the above

5) Which of the following machine learning techniques helps in detecting the outliers in data?
a) Clustering
b) Classification
c) Anamoly detection
d) All of the above
Answer is C
6) Identify the incorrect numerical functions in the various function representation of machine
learning.
a) Support Vector
b) Regression
c) Case based
d) Classification
Answer is C
7) Analysis of ML algorithm needs
a) Statistical learning theory
b) Computational learning theory
c) None of the above
d) Both a and b
Answer is D
8) Identify the difficulties with the k-nearest neighbor algorithm.
a) Curse of dimensionality
b) Calculate the distance of test case for all training cases
c) Both a and b
d) None
Answer is C

9) The total types of the layer in radial basis function neural networks is
a) 1
b) 2
c) 3
d) 4
Answer is C
10) Which of the following is not a supervised learning
a) PCA
b) Naïve bayes
c) Linear regression
d) KMeans
Answer is A
Allower is A
11) What is unsupervised learning?
a) Number of groups may be known
b) Features of groups explicitly stated
c) Neither feature nor number of groups is known
d) None of the above
Answer is C
12) Which of the following is not a machine learning algorithm?
a) SVM
b) SVG
c) Random Forest Algorithm
d) None of the above
Answer is B
13) is the scenario when the model fails to decipher the underlying trend in the input data
a) Overfitting

b) Underfitting
c) Both a and b
d) None of the above
Answer is B
14) Real-Time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are
applications of
a) Reinforcement learning
b) Supervised learning
c) Unsupervised Learning
d) None of the above
Answer is A
15) What is called the average squared difference between classifier predicted output and actual
output?
a) Mean relative error
b) Mean squared error
c) Mean absolute error
d) Root mean squared error
Answer is B
<b>16)</b> Logistic regression is a regression technique that is used to model data having a
outcome.
a) Linear, binary
b) Linear, numeric
c) Nonlinear, binary
d) Nonlinear, numeric
Answer is C
17) You are given reviews of few netflix series marked as positive, negative and neutral. Classifying

reviews of a new netflix series is an example of



- (C) Underfitting
- (D) SVM is highly flexible

#### Answer is A

- 22) Increase in which of the following hyper parameter results into overfit in Random forest?
- (1). Number of Trees. (2). Depth of Tree, (3). Learning Rate
- (A) Only 1
- (B) Only 2
- (C) 2 and 3
- (D) 1,2 and 3

#### Answer is B

**23)** Below are the 8 actual values of target variable in the train file: [0,0,0, 0, 1, 1,1,1,1,1], What is the entropy of the target variable?

# $(A) - (6/10 \log(6/10) + 4/10 \log(4/10))$

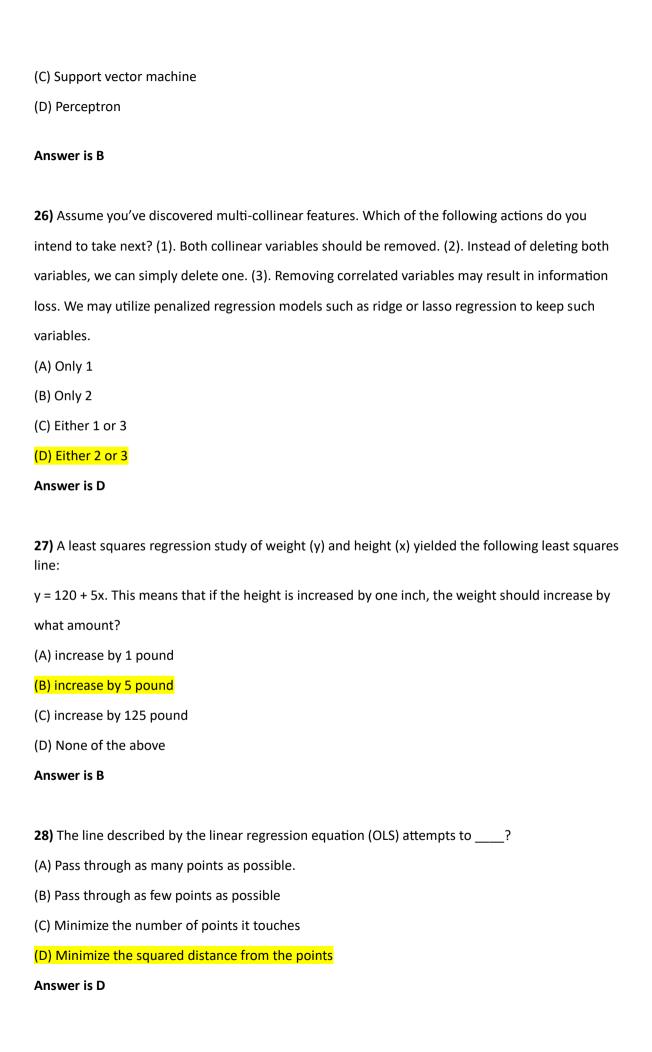
- (B)  $6/10 \log(6/10) + 4/10 \log(4/10)$
- (C)  $4/10 \log(6/10) + 6/10 \log(4/10)$
- (D)  $6/10 \log(4/10) 4/10 \log(6/10)$

## Answer is A

- 24) Lasso can be interpreted as least-squares linear regression where
- (A) weights are regularized with the I1 norm
- (B) weights are regularized with the I2 norm
- (C) the solution algorithm is simpler

## Answer is A

- **25)** Consider the problem of binary classification. Assume I trained a model on a linearly separable training set, and now I have a new labeled data point that the model properly categorized and is far away from the decision border. In which instances is the learnt decision boundary likely to change if I now add this additional point to my previous training set and re-train? When the training model is,
- (A) Perceptron and logistic regression
- (B) Logistic regression and Gaussian discriminant analysis



- 29) For two real-valued attributes, the correlation coefficient is 0.85. What does this value indicate?
- (A) The attributes are not linearly related
- (B) As the value of one attribute increases the value of the second attribute also increases
- (C) As the value of one attribute decreases the value of the second attribute increases
- (D) The attributes show a curvilinear relationship

### Answer is C

- **30)** Which neural network architecture would be most suited to handle an image identification problem (recognizing a dog in a photo)?
- (A) Multi Layer Perceptron
- (B) Convolutional Neural Network
- (C) Recurrent Neural network
- (D) Perceptron

Answer is B