BUDDI Machine Learning Quiz

You have 60 minutes to complete the quiz.

There are 61 questions in total and this is the split up:

Machine Learning (32 questions)
Deep Learning (15 questions)
Image Processing (6 questions)
Natural Language Processing (8 questions)

All the best,
- Team BUDDI

Candidate Basic Info

1

Name *

Harsh Gupta

2

Email *

harshgupta57100@gmail.com

3

Contact No. *

9326224289

Machine Learning (32 questions)

4

Suppose you are interested in finding a parsimonious model (simpler model with the desired level of performance) to predict housing prices?
(2 Points)

- Lasso (L1) Regression
- Ordinary Least Squares Regression
- Logistic Regression
- Ridge (L2) Regression

5

When batch gradient descent is performed, what is the expected distribution (theoretically) of batch losses for an epoch?
(3 Points)

- Negative Exponential Distribution
- Uniform Distribution
- Gaussian Distribution
- Poission Distribution

6

Which of the following statements about the k-means algorithm are correct?
(1 Point)

- [x] The k-means algorithm is sensitive to outliers.
- [ ] K-means algorithm estimates stable and repeatable clusters
- [x] The centroids may not be true data points.
- [ ] The k-means algorithm can directly handle categorical data.

7

Consider building a classifier for detecting fraudulent credit card transactions. Which among the following prefer?
(1 Point)

- ( ) Aim for high precision on fraudulent class
- ( ) Aim for high precision on non-fraudulent class
- ( ) Aim for high recall on non-fraudulent class
- (•) Aim for high recall on fraudulent class

8

There are 'n' bins of which the kth bin contains (k-1) blue balls and (n-k) red balls.
You pick a bin at random and remove two balls at random without replacement. Find the probability that:
1) the second ball is red;
2) the second ball is red, given that the first is red
(2 Points)

- (●) 1/2, 2/3
- ( ) 1/3, 2/3
- ( ) 1/2, 1/3
- ( ) 1/3, 1/3

9

In a binary class imbalance scenario, which is the best performance metric? (1 Point)

- (●) Precision and Recall
- ( ) ROC
- ( ) Weighted F1
- ( ) Accuracy

10

In a multiclass classification, which is the best performance metric? (2 Points)

- [ ] Macroaveraged F1 when the best performance is expected only in classes with high samples.
- [ ] Microaveraged F1 when the best performance expected among all classes.
- [x] Macroaveraged F1 when the best performance expected among all classes.
- [x] Microaveraged F1 when the best performance is expected only in classes with high samples.

11

How can we classify a non-linearly separable set of data points using a large margin classifier like SVM?
(1 Point)

- Remove non-linear data points
- Kernel does not matter
- Polynomial kernel
- Linear kernel

12

When the determinant of a matrix is close to ZERO, which of the following is incorrect?
(1 Point)

- Matrix is Invertible
- Matrix is not Invertible
- Matrix contains dependent columns or rows
- Matrix is diagonal

13

After fitting a linear regression to time-series data, you find the residual plot to be non-linear What is your conclusion?
(3 Points)

- [x] K-NN Regression might fit the data better.
- [ ] Linear time series models suffice the data

- [ ] Non - Random plots are a consequence of the central limit theorem.
- [x] Non-linear models might fit the data better.

14

Which of the following supervised machine learning methods are NOT affected by feature scaling?
(1 Point)

- ( ) Neural Networks
- ( ) KNN
- ( ) Naive Bayes
- (•) Decision Trees

15

A medical company touts its new test for a certain genetic disorder. The false negative rate is small: if you have the disorder, the probability that the test returns a positive result is 0.999. The false positive rate is also small: if you do not have the disorder, the probability that the test returns a positive result is only 0.005. Assume that 2% of the population has the disorder. If a person chosen uniformly from the population is tested and the result comes back positive, what is the probability that the person has the disorder?
(2 Points)

- ( ) 0.02
- ( ) 0.204
- ( ) 0.976
- (•) 0.803

16

Which of the following tasks can be "best" solved using clustering?
(1 Point)

- (●) Training a robot to solve a maze
- ( ) Stock market price prediction
- ( ) Predicting the amount of rainfall based on various cues
- ( ) Detecting fraudulent credit card transactions

17

You are given the following five training instances, x = [(2,1); (6,3); (2,5); (6,7); (10,7)] and y = [4; 2; 2; 3; 3]. We want to model this data set using the K-nearest neighbor regressor model. What would be the value of y corresponding to a test point t = [(3,6)]?
(2 Points)

- [ ] For K = 2, y = 3
- [x] For K = 2, y = 2.5
- [x] For K = 3, y = 2.33
- [ ] For K = 3, y = 2.666

18

Which of the following is true about Ridge(L2) regression?
(3 Points)

- [x] When lambda goes to infinity, we get very small coefficients approaching zero.
- [x] When lambda is 0, model works like linear regression model

- [ ] When lambda is 0, model doesn't work like linear regression model
- [ ] When lambda goes to infinity, we get very large coefficients.

19

An airline knows that 5 percent of the people making reservations on a certain flight will not show up. Consequently, their policy is to sell 52 tickets for a flight that can hold only 50 passengers. What is the probability that there will be a seat available for every passenger who shows up?
(2 Points)

- ( ) 0.5101
- ( ) 0.6308
- ( ) 0.81
- (•) 0.7405

20

Consider building a spam filter for distinguishing between genuine e-mails and unwanted spam e-mails. Assuming spam to be a positive class, which among the following would be more important to optimize?
(1 Point)

- (•) Precision
- ( ) Recall
- ( ) Accuracy

21

Consider a modified k-NN method in which once the k nearest neighbors to the query point are identified, you do a linear regression fit on them and output the fitted value for the query point. Which of the following is/are true regarding this method.
(3 Points)

- [x] This method makes an assumption that the data is locally linear.
- [x] This method has higher variance compared to K-NN
- [ ] This method has higher bias compared to K-NN
- [x] In order to perform well, this method would need dense distributed training data.

22

In a medical study, doctors record the average calorie intake for a group of adolescents in a specific year and their corresponding average increase in height for the same period. If the doctors are to perform regression analysis on this data, which variable should they consider to be the independent variable and which one should they consider to be the dependent variable?
(2 Points)

- ( ) Independent: Increase in height; Dependent: Calorie intake
- (•) Independent: Calorie intake; Dependent: Increase in height

23

In building a linear regression model for a particular data set, you observe the coefficient of one of the features having a relatively high negative value. This suggests that,
(2 Points)

- (•) It is not possible to comment on the importance of this feature without additional information

- ○ This feature has a strong effect on the model (should be retained)
- ○ This feature does not have a strong effect on the model (should be ignored)

24

Choose the correct option(s) from the following
(3 Points)

- ☐ When working with a small dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.
- ☑ When working with a large dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.
- ☑ When working with a small dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.
- ☐ When working with a large dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.

25

Which of the following properties are NOT characteristic of fully grown decision trees?
(2 Points)

- ○ Lack of smoothness of decision surface
- ○ Non-parametric modeling
- ◉ High bias
- ○ High variance

26

Which among the following are some of the differences between bagging and boosting?
(3 Points)

- [x] In comparison with the performance of a base classifier on a particular data set, bagging will generally not increase the error whereas as boosting may lead to an increase in the error
- [ ] In bagging we use the same classification algorithm for training on each sample of the data, whereas in boosting, we use different classification algorithms on the different training data samples
- [x] In bagging we typically use sampling with replacement whereas in boosting, we typically use weighted sampling techniques

27

Which of the following are true for any A(m*n) matrix of real numbers.
(2 Points)

- ( ) The rowspace of A is the same as the columnspace of A
- ( ) The rowspace of A is the same as the rowspace of A'
- ( ) The eigenvectors of AA' are the same as the eigenvectors of A' A
- (•) The eigenvalues of AA' are the same as the eigenvalues of A'A

28

Given that we can select the same feature multiple times during the recursive partitioning of the input space, is it always possible to achieve 100% accuracy on the training data, given that we allow for trees to grow to their maximum size? (Hint: Think noise)
(1 Point)

- ( ) Yes

- (●) No

29

Which of the following statements is true?
(1 Point)

- ( ) T-test is used when the number of samples is large
- (●) T-test is used when the number of samples is small.
- ( ) T-test assumes the underlying distribution is a beta distribution.
- ( ) T-test assumes the underlying distribution is a exponential distribution.

30

Suppose you are an online advertiser (like Google Ads), which accepts advertisements (consisting of short text and a link) from your customers (companies, such as say Samsung or Hindustan Unilever). You need to build a system, which on submitting an ad-page to it, classifies it as spam or not spam, and immediately adds it to your corpus of ads if it is not spam. Your development team has come up with two systems - system A and system B, to perform this task. You need to evaluate which system is better for the task using hypothesis testing based methods. Which of these variables are likely to be extraneous to the task? (Note that multiple answers may be correct)
(1 Point)

- [ ] Regional market in which the system will be deployed (India, Canada or USA)
- [x] Average Click-Through Rate (fraction of users who open the link in the ad) for ads which you classify as non-spam.
- [x] Classification Accuracy of the system
- [ ] Month of the year

31

Which among the following is/are some of the assumptions made by the k-means algorithm (assuming Euclidean distance measure)?
(1 Point)

- [ ] Clusters are elliptical in shape
- [x] There is no wide variation in density among the data points
- [x] Data points in one cluster are well separated from data points of other clusters
- [ ] Clusters are of dissimilar radii

32

Which among the following prevents overfitting when we perform bagging?
(1 Point)

- [ ] The practice of validation performed on every classifier trained
- [x] The use of weak classifiers
- [ ] The use of classification algorithms which are not prone to overfitting
- [ ] The use of sampling with replacement as the sampling technique

33

Suppose we are given 'n' p-dimensional data points and the corresponding class labels (k different classes). We want to build a decision tree classifier to classify the data. However, we find that there are missing values in the data set. Is it possible to use the EM algorithm to fill the missing data given the above information and make no further assumptions?
(3 Points)

- ( ) Yes

- ○ Maybe
- ◉ No

34

Which of the following holds true with respect to Principal Component Analysis (PCA)?
(2 Points)

- ☑ PCA can approximate non linear manifolds during dimensionality reduction
- ☑ All principal components are mutually orthogonal to each other
- ☑ Direction of the eigen vectors calculated represent the direction of most variance
- ☑ It is an unsupervised learning algorithm.

35

Consider a scenario where you don't know the number of clusters for a given clustering problem, which of the following clustering algorithm would you prefer?
(1 Point)

- ○ KMeans clustering
- ○ Spectral clustering
- ◉ Heirarchial clustering
- ○ Density Based clustering

Deep Learning (15 questions)

36

An image shape is 227x227x3, Slide=4, Filter=96x96 and Padding=1. Find the output shape of the image after applying the filter?
(2 Points)

- (●) 34
- ( ) 33
- ( ) 30
- ( ) 36

37

Sathya used dropout methodology in a Deep Neural Network(DNN) with a dropout rate = P% for the i-th layer.

Which of the following is true for the Train phase?
(1 Point)

- ( ) P% weight will only be considered from whole DNN while training.
- ( ) P% weight will only be considered from whole DNN's i-th layer while training.
- ( ) P% nodes from whole DNN will not be considered while training.
- (●) P% nodes from whole DNN's i-th layer will not be considered while training.

38

With respect to back propagation algorithm of deep neural networks, which of the following is correct?
(3 Points)

- [✓] It is also called generalized delta rule

- [x] Error in output is propagated backwards only to determine weight updates
- [x] There is feedback of signal at every stage
- [ ] None of the above

39

Which of the following is a learning(optimization) algorithm?
(1 Point)

- [x] Gradient Descent
- [x] Adam
- [ ] Logistic
- [x] RMS Prop

40

Which of the following are true about the Perceptron classifier?
(2 Points)

- [x] It can learn a OR function
- [ ] It can learn a XOR function
- [x] Hyperplane does not depend on the order in which the points are presented to the training process
- [x] If linearly non-separable, there exists some initial weights which might lead to non-convergence

41

Assume that you are given a data set and a neural network model trained on the data set. You are asked to build a decision tree model with the sole purpose of understanding/interpreting the built neural network model. In such a scenario, which among the following measures would you concentrate most for optimization?
(3 Points)

- ○ Accuracy of the decision tree model on the given data set
- ○ F1 measure of the decision tree model on the given data set
- ◉ Fidelity (the fraction of instances that both neural network and the decision tree concur)
- ○ Comprehensibility (measured in terms of the size of the corresponding rule set)

42

Which of the following statements is/are true about Neural Networks?
(2 Points)

- ☑ Neural Networks can model arbitrarily complex decision boundaries.
- ☑ Neural Networks can be used to emulate a Gaussian kernel SVM
- ☑ Training of a neural network is very sensitive to the initial weights.
- ☐ Ideal initialization for weights would be setting all of them to zeros

43

Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5x5. How many parameters does this hidden layer have (including the bias parameters)?
(2 Points)

- ○ 2501

- 2600
- 7500
- 7600

44

Assert the statement below.
Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation.
(2 Points)

- True
- False
- Partially

45

The hyperbolic-tan activation usually works better than sigmoid activation function for hidden units
(3 Points)

- True, because the mean of its output is closer to zero, and so it centers the data better for the next layer
- True, because the mean of its output is closer to one, and so it centers the data better for the next layer
- False, because the mean of its output is closer to zero, and so it centers the data better for the next layer
- False, because the mean doesn't affect the activation function

46

Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements are true?
(1 Point)

- Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
- Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have broken symmetry
- Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished symmetry breaking as described in lecture.
- The first hidden layers neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way

47

A common method to accelerate the training of Generative Adversarial Networks (GANs) is to update the Generator k times for every 1 time you update the Discriminator.
(1 Point)

- True
- False
- It depends on the architecture of the GAN

48

You want to map every possible image of size 64 x 64 to a binary category (cat or non-cat). Each image has 3 channels and each pixel in each channel can take an integer value between (and including) 0 and 255. How many bits do you need to represent this mapping?
(2 Points)

- 256^(3^(64 x 64))
- 256^(3 x 64 x 64)
- (64 x 64)^(256 x 3)
- (256 x 3)^(64 x 64)

49

Which of the following are techniques that could be used to mitigate overfitting in neural networks?
(2 Points)

- [x] Enforcing weight and kernel constraints for neurons.
- [x] Using Dropout to randomly deactivate input neurons
- [ ] Increasing the number of trainable parameters of the model
- [ ] All of the above.

50

Which of the following are true with respect to Transfer Learning?
(2 Points)

- [x] It involves the usage of a network pre-trained on a prior/previous dataset to extract features from a new dataset

- [ ] Transfer learning involves the projection of weights into a different hyperplane for better learning purposes
- [x] Transfer Learning is of benefit for training on datasets that have significantly less data.
- [ ] All of the above

Image Processing (6 questions)

51

Which of the color space representation listed below, best emulate human visual perception?
(1 Point)

- (•) RGB
- ( ) HSV
- ( ) CbCr
- ( ) CMYK

52

Which of the following are reasons against using raw pixel values of images as features with a multi-layer perceptron?
(1 Point)

- [x] Loss of structural context for images
- [ ] Requires larger neural network
- [x] Pixels are translation, illumination and scale variant
- [ ] None of the above

53

Interpolation techniques find widespread usage in image processing, such as resizing and demosaicing. Which of the following interpolation techniques would produce images of relatively sharper quality?
(1 Point)

- Nearest Neighbour Interpolation
- Mean-based interpolation
- Bilinear Interpolation
- Bicubic interpolation

54

Which of the following activation function can't be used at the output layer to classify an image?
(1 Point)

- Sigmoid
- Softmax
- ReLU
- None of the above

55

Which of the following metrics are used to quantify the performance of algorithms used for image restoration
(1 Point)

- Peak Signal To Noise Ratio
- Degradation Index

- Jaccard Similarity index
- Structural Similarity Index Measure

56

Sobel filter is not ideal for detection of which of the following?
(1 Point)

- Horizontal lines
- Vertical lines
- Diagonal lines
- Edges

Natural Language Processing (8 questions)

57

1. Latent Dirichlet Allocation
2. Latent Semantic Indexing
3. Keyword Normalization

Given a term-document matrix of the input data of 100K documents, which of the following remediation can be used to reduce the data dimensions?
(1 Point)

- 2 only
- All of them
- 2 and 3 only
- 1 only

58

You have created a document term matrix of the data, treating every tweet as one document. Which of the following statements are correct?
(1 Point)

- Matrix is Dense
- Matrix is Diagonal
- Matrix is Singular
- Matrix is Sparse

59

Which of the statements are correct with respect to Topic Modeling?
(1 Point)

- LDA (Linear Discriminant Analysis) can be used to perform topic modeling
- It is a supervised learning technique
- Topic count does not depend on the size of data
- Feature are assumed to nominal

60

Which of the following techniques can be used for the purpose of keyword normalization, the process of converting a keyword into its base form?
(1 Point)

- Stemming
- Levenshtein
- Soundex

- Lemmatization

61

While working with context extraction from text data, you encountered two different sentences: (1) The tank is full of soldiers. (2) The tank is full of nitrogen. Which of the following measures address the problem of word sense disambiguation?
(1 Point)

- Use knowledge bases and Ontologies
- Use dependency parsing of sentence to understand the meanings
- Co-reference resolution in which one resolute the meaning of ambiguous word with the proper noun present in the previous sentence
- Compare the dictionary definition of an ambiguous word with the terms contained in its neighborhood

62

Which of the following statements are true for the Word2Vec model?
(1 Point)

- All of the Above
- CBOW and Skip-gram are shallow neural models
- CBOW is a Recurrent Neural Network model
- Word2Vec is a probabilistic language model

63

Which of the following models can be used for the purpose of document similarity?
(1 Point)

- Latent Semantic Analysis model
- word2vec model
- All of the above
- Linear Discriminant Analysis model

64

Google's "Did you mean" feature is a mixture of different techniques. Which of the following techniques are likely to be ingredients?
(1 Point)

- Sequence modeling
- Levenshtein distance against the dictionary terms
- Collaborative Filtering
- Translation of sentences into multiple languages