Practical Machine Learning Practice Session

Q1. Among the following option identify the one which is not a type of learning.

    a) Semi unsupervised learning
    b) Supervised learning
    c) Reinforcement learning
    d) Unsupervised learning

Q2. You've just finished training a random forest for spam email classification, and it gives abnormally bad performance on your validation set, but good performance on your training set. What could be causing the problem? Pick the most appropriate option.

    a) Your decision trees are too deep.
    b) You are randomly sampling too many features to consider when you choose a split.
    c) You have too few trees in your ensemble.
    d) All the above

Q3. After applying a regularization penalty in linear regression, you find that some of the coefficients of w are zeroed out. Which of the following penalties might have been used?

    a) L1 norm
    b) L2 norm
    c) Both A and B.
    d) None of the Above.

Q4. We can't use the following techniques to represent text documents as vectors:

    a) Binary BOW.
    b) TF IDF weighted BOW.
    c) Average word2vec
    d) None

Q5. Application of machine learning methods to large databases is called
    a) data mining.
    b) artificial intelligence
    c) big data computing
    d) internet of things

Q6. If machine learning model output involves target variable, then that model is called as

    a) descriptive model

b) predictive model
c) reinforcement learning
d) all the above

Q7. In what type of learning labelled training data is used

a) unsupervised learning
b) supervised learning
c) reinforcement learning
d) active learning

Q8. In following type of feature selection method, we start with empty feature set
a) forward feature selection
b) backword feature selection
c) both a and b??
d) none of the above

Q9. In PCA the number of input dimension are equal to principal components

a) True
b) False

Q10. PCA can be used for projecting and visualizing data in lower dimensions.

a) True
b) False

Q11. 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
a) supervised learning
b) unsupervised learning
c) semi supervised learning
d) reinforcement learning

Q12. A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is

a) nominal
b) ordinal
c) categorical
d) Boolean

Q13. Supervised learning and unsupervised clustering both require which is correct according to the statement.

a) output attribute.
b) hidden attribute.
c) input attribute.
d) categorical attribute

Q14. The "curse of dimensionality" refers:
   a) all the problems that arise when working with data in the higher dimensions, that did not exist in the lower dimensions.
   b) all the problems that arise when working with data in the lower dimensions, that did not exist in the higher dimensions.
   c) all the problems that arise when working with data in the lower dimensions, that did not exist in the lower dimensions.
   d) all the problems that arise when working with data in the higher dimensions, that did not exist in the higher dimensions.

Q15. Feature can be used as a
   a) binary split
   b) predictor
   c) both a and b??
   d) none of the above

Q16. Support Vector Machine is
   a) logical model
   b) proababilistic model
   c) geometric model
   d) none of the above

Q17. A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.
   a) true
   b) false
   c) sometimes – it can also output intermediate values as well

Q18. How can SVM be classified?
   A. it is a model trained using unsupervised learning. it can be used for classification and regression.
   B. it is a model trained using unsupervised learning. it can be used for classification but not for regression.
   C. it is a model trained using supervised learning. it can be used for classification and regression.
   D. t is a model trained using unsupervised learning. it can be used for classification but not for regression.

Q19. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

 a) the model would consider even far away points from hyperplane for modelling
 b) the model would consider only the points close to the hyperplane for modelling
 c) the model would not be affected by distance of points from hyperplane for modelling
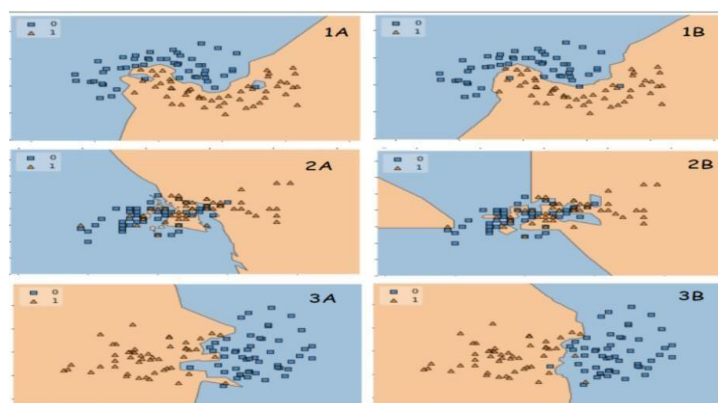 d) none of the above

Q20. Which of the following methods cannot achieve zero training error on any linearly separable dataset?

 a) decision tree
 b) 15-nearest neighbours
 c) hard margin SVM
 d) perceptron

Q21. In a weighted K-NN, we can make class predictions by these following approaches:

 a) We compute the class-wise sum of the distances of the neighbours from the query point and then assign the class with the largest sum-ofdistances value.
 b) We compute the class-wise sum of the inverse of the distances of the neighbours from the query point and then assign the class with the lowest sum value.
 c) We compute the class-wise sum of the distances of the neighbours from the query point and then assign the class with the lowest sum value.
 d) None

Q22. Which of the images among the following pairs represents overfitting?

 a) 1B, 2B, 3A
 b) 1A, 2B, 3A
 c) 1A, 2A, 3A
 d) D None of the above

Q23. Which of the following statements are true for KNNs?

a) The complexity of the decision surface increases with increase in value of k.
b) Manhattan distance cannot be used as a distance metric with k-NN.
c) The space complexity of KNN is the same as it's time complexity.
d) None of the above

Q24. Given that the euclidean distance between 2-unit vectors x1 & x2 is 1.075 units, the angle between them is____. You can use a calculator to solve this problem.
a) Approx 45°
b) Approx 70°
c) Approx 65°
d) None of the above.

Q25. In a corpus of n documents, one document is randomly picked. The document contains a total of t terms and the term "performance" appears k times. What is the TF-IDF, of the term "performance" if it appears in approximately 25% of the total documents?
a) kt * Log (0.25)
b) k * Log (4) / t
c) Log (4) / kt
d) t * Log (0.25) / k

Q26. How many unique features will be generated if we use ngram_range = (1, 2) as the parameter in scikit-learn for Bag-of-Words representation of the sentence: "Wikipedia is the place to start your search about the subject"
a) 15
b) 8
c) 20
d) 10

Q27. Consider the following corpus:
"This product is great"
"Please do not buy"
"This is worth it"
"One of a kind"
The bag of words vectors for the above corpus will have a dimensionality of: a) 4
b) 16
c) 14
d) None of the above

Q28. Consider the ξ_i (=slack) values in a Soft-margin SVM formulation for points marked 1, 2, 3 as shown in the image below. What is the correct order for these slack variables? Here f
= w.x+b and blue points are positive labelled points and yellow ones are negative labelled points.
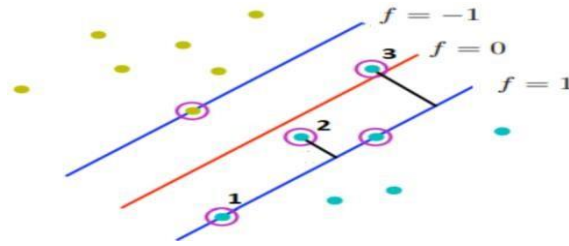
a) ξ1 > ξ2 > ξ3
b) ξ1 < ξ2 > ξ3
c) ξ1 < ξ2< ξ3
d) None of the above

Q29. Consider a uniform discrete distribution for a random variable K which can take up to P
different values {n1, n2, n3, np......}. What will be the entropy value for the random variable K?

a) P logP
b) -logP
c) 1
d) logP

Q30. Let us say a logistic regression model with no bias term is fit to a dataset with 4 features x1, x2, x3 and x4 and binary response variable, y. The weights w1, w2, w3, w4 found corresponding to features x1, x2, x3,x4 respectively follow the below conditions:

|w3|>|w2|>|w1|>|w0|
w1>w0>w2>w3

Given a new datapoint = <1,1,1,1>. Which class does it belong to assuming the threshold of 0.5 probability?

a) Positive class
b) Negative class
c) Not enough information
d) Sometimes positive and sometimes negative

Q31. Given that there are 100 positive points in the dataset.
Say we have 2 models M1 & M2.
Model M1 predicts just one point as positive and it is a correct prediction based on ground truth values.
Model M2 predicts 300 points as positive, and within these 300 points all the groundtruth positive points are present. Then, which of the following options are correct.

a) Precision of M1 is 0.01
b) Recall of M2 is 1
c) Recall of M2 is 0.33

d) None of the above

Q32. Feel free to use a calculator for this question.
Consider the following dataset for regression:

	0	1	2	3	4	5	6	7	8	9
X	0.0	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00
y	1.0	3.43	2.33	1.96	4.05	4.72	5.85	4.77	4.75	4.47

Given that a line with slope = 0.35 & intercept = 2 is found to be the best fit line after performing regression, the coefficient of determination (R^2 metric) is:
- a) Approx 0.45
- b) Approx 0.85
- c) Approx 0.5
- d) Approx 0.6

Q33. Given the values of x and the corresponding y values as shown below:

x	y
0	3
1	4
2	6

If the equation of the linear regression line fit on this data is y= ax+b. What are the values of a and b (approximately)? Feel free to use a calculator, but not a code editor or Python-Notebook.
- a) 1.0,3.12
- b) 1.5,2.83
- c) 1.7,2.34
- d) 2.1,1.92

Q34. Match the following gradient optimization techniques with their process/method of optimization:

a) Gradient descent   1) processes all the training examples for each iteration of gradient descent

b) Stochastic gradient descent   2) processes 1 training example per iteration.

c) Mini batch gradient descent   3) A subset of the training data is processes per Iteration

- a) (a, 1), (b, 2), (c, 3)
- b) (a, 3), (b, 1), (c, 2)
- c) (a, 1), (b, 3), (c, 2)
- d) None of the above

Q35. What is the space complexity of logistic regression?
- a) $O(d)$
- b) $O(d2)$

c) O(dlogd)
d) O(logd)

Q36. Which of the following is true about bagging in random forests?
1. Bagging can be done in parallel
2. We sample the data without repetition for each tree.
3. Bagging helps in reducing overfitting
   a) 1 and 2
   b) 2 and 3
   c) 1 and 3
   d) All of these

Q37. Feature standardization is not required in decision trees:
   a) Correct
   b) Incorrect
   c) It is mandatory to do it if the data is not linearly separable
   d) It is mandatory to do it if we are dealing with data that contains outliers

Q38. In TF IDF which of the following are true:
   a) TF IDF = TF + IDF
   b) As the number of repetitions of a word in the corpus increases TF decreases
   c) As the number of repetitions of a word in the sentence increases IDF decreases
   d) None of the above

Q39. With reference to the assumptions of the linear regression, which statement is not correct?
   a) Linear Relationship between X and mean of Y
   b) Observations are independent of each other
   c) For any fixed value of X, Y is normally distributed
   d) None of the above

Q40. What is the domain of sigmoid function?
   a) -1 to 1
   b) 0 to 1
   c) -infinity to infinity
   d) 0 to infinity

Q41 The maximum and minimum entropy for a two-class dataset is:
   a) Maximum = 1, Minimum = -1
   b) Maximum = 1, Minimum = 0
   c) Maximum = 0.5, Minimum = -1
   d) Maximum = 0.5, Minimum = 0

Q42 Given the following information about a model's performance:
F1 score = 0.8
Recall = 0.8

   a) The precision of the model will be:
   b) .75
   c) .8
   d) .85
   e) None of the above

Q43 Consider the data points and their corresponding target variables shown below:

x1	x2	y
0.55	-0.3	0
-0.05	-1.95	1
-0.77	0.29	1
0.54	0.78	0

Given that the best logistic regression equation is of the form y=
sigmoid($\mathbf{w}.x+\mathbf{w0}$), where w = [-0.95, -0.3])
What is the best value of '$\mathbf{w0}$' among the 4 values given below?
   a) -0.22
   b) 0.15
   c) 0.33
   d) -0.09

Q44 Given a confusion matrix of the predictions of a classifier for a multiclass
classification problem. For the model to be considered reasonable, which of the
following should be true.
   a) The values of its principal diagonal elements must be high.
   b) The values of its off-diagonal elements must be high
   c) Both principal and off diagonal elements must be high
   d) None of the above

Q45 Given that the training data has outliers and say, we train a KNN classification
model on it. After cross validation, we observe the following accuracy scores for various
values of K: K=1: accuracy = 0.95
K=3: accuracy = 0.95
K=5: accuracy = 0.95
K=7: accuracy = 0.93
K=8: accuracy = 0.90
Which value of K would you be the most appropriate to use:
   a) K=1
   b) K=2
   c) K=5
   d) K=7

Q46 Which of the following models can produce feature importance scores as part of the modelling process.
- a) KNN
- b) Logistic Regression
- c) Kernel-SVM
- d) None of these

Q 47. Which of the following equations is correct?
- a) P(A, B, C) = P(A|B, C) P(B|C) P(C)
- b) P(A|B, C) = P(A, B, C) P(B|C) P(B)
- c) P(A, B, C) = P(B|A, C) P(C|A) P(B)
- d) None of the above

Q48. Which of the following holds true for logistic regression?
- a) Logistic regression is basically Gaussian Naive Bayes with Poisson distribution assumption for the class labels.
- b) Logistic regression is basically Gaussian Naive Bayes with Bernoulli distribution assumption for the class labels.
- c) Logistic regression is basically Gaussian Naive Bayes with exponential distribution assumption for the class labels.
- d) None of the above

Q49 Given 4 vectors x1, x2, x3 & x4, such that:
They lie on a unit circle at the origin in the same clockwise order as mentioned above.
The euclidean distance between x1 & x2 = 1.075 units.
The euclidean distance between x2 & x3 = 1 unit.
x3 is perpendicular to x4
What is the approximate (rounded to one decimal place) euclidean distance between x1 and x4?
- a) 3.4
- b) 2.5
- c) 1.9
- d) None of the above

Q50 Consider the data points and their corresponding target variables shown below:

	f1	f2	y
1	-1.54	-0.41	0
2	-0.02	0.2	1
3	-0.19	0.6	1

Given that the best logistic regression equation is of the form y= sigmoid(w.x+w0), where w = [-0.52, w2] and w0 = 0.45
What is the best value of 'w2' among the 4 values given below?

a) 1
b) -1
c) 2
d) -2

Q51 The number of weights that get updated during back propagation in max pooling layer with input of size m*m*c and kernel size n*n*c?

a) Zero
b) (n*n*c)
c) (n*n*c+1)
d) None of the above

Q 52. Given Input image shape = (128, 128), kernel_size= (5,5), stride=1 & padding= 0, What is the output size of the convolution operation?

a) (128, 128)
b) (124, 124)
c) (125, 125)
d) (123, 123)

Q53. Which of the following pre-trained models in Pytorch can be fine-tuned for image classification?

a) VGG16
b) ResNet
c) InceptionNet
d) All the above

Q54. The number of times the cell state of an LSTM is modified during one recurrent step is:

a) One
b) Two
c) Three
d) None of Above

Q55. Which among the RNN configurations given below is appropriate for the task of Time Series forecasting:
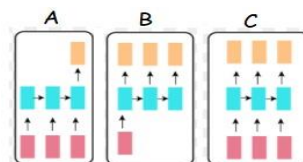
a) A,B & C
b) A&B
c) A&C
d) C

Q56. Given a kernel of w x h = n , what are the number of parameters that we are learning in a CNN layer which: Outputs 3 feature maps and Receives an input of size is m*m*c.

a) (m*m*c)
b) ((n*n*c)+1) x 3
c) (m*n*c) x 3
d) ((m*m*c)+1) x 3

Q57 Given an LSTM, the output cell state will be the same as the input cell state if:
The outputs from the forget & input gates are unit vectors
The outputs from the forget & input gates are zero vectors
The outputs from the forget & input gates are a unit vector & a zero vector respectively
The outputs from the forget & input gates are a zero vector & a unit vector respectively

Q58 What are the number of multiplication operations required (approx) when we use a convolution layer with 32 kernels of size 5x5 on an input of shape 28x28x192, using stride 1 and 'same' padding.

a) 120 million
b) 60 million
c) 4 million
d) 11 million

Q 59. Given the image of the LSTM shown below, Which of the following statements are correct:
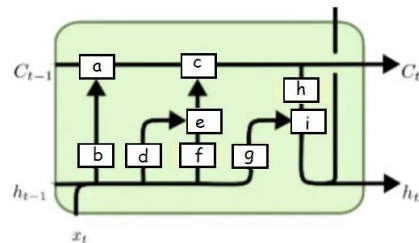
a)
a represents multiplication of 2 vectors
b) c represents addition of 2 vectors
c) e represents multiplication of 2 vectors
d) None

Q60. Bidirectional RNNs can be used for which of the following tasks:

a) Spontaneous speech to text conversion
b) Time Series forecasting
c) Text classification
d) None of the above

Q61 Given that we pass an input image of 415x415x3 through YOLO v3 model's backbone network, the size of the tensor outputted at the last layer is

a) 32x32x512
b) 13x13x1024

c) 26x26x1024
d) 32x32x1024

Q62. Assume we have 64 classes, and the size of the output feature map is 14*14. The total number of values present in the output tensor of YOLO-V3 is 94668 Choose the correct option.

a) Number of anchor boxes is 2.
b) Number of anchor boxes is 3.
c) Number of anchor boxes is 7.
d) None

Q63 What do you mean by generalization error in terms of the SVM?

a) how far the hyperplane is from the support vectors
b) how accurately the svm can predict outcomes for unseen data
c) the threshold amount of error in an svm

Q64. What is the purpose of performing cross- validation?

a) a. to assess the predictive performance of the models
b) b. to judge how the trained model performs outside the
c) c. both a and b

Q65 Techniques involve the usage of both labelled and unlabelled data is called   .

a) supervised
b) semi- supervised
c) unsupervised
d) none of the above

Q66 Another name for an output attribute.

a) predictive variable
b) independent variable
c) estimated variable
d) dependent variable

Q67 if there is only a discrete number of possible outcomes (called categories), the process becomes a_____.

a) Regression
b) Classification.
c) Modelfree
d) Categories

Q68. What is 'Overfitting' in Machine learning?

a) when a statistical model describes random error or noise instead of underlying relationship overfitting' occurs.
b) Robots are programed so that they can perform the task based on data they gather from sensors.

c) While involving the process of learning 'overfitting' occurs.
d) a set of data is used to discover the potentially predictive relationship

Q69. What is 'Test set'?
a) Test set is used to test the accuracy of the hypotheses generated by the learner.
b) It is a set of data is used to discover the potentially predictive relationship.
c) Both A & B
d) None of above

Q70. If Linear regression model perfectly first i.e., train error is zero, then _____
a) Test error is also always zero
b) Test error is nonzero
c) Couldn't comment on Test error
d) Test error is equal to Train error

Q71 Naive Bayes classifiers are a collection ------------------of algorithms
a) Classification
b) Clustering
c) Regression
d) All

Q72 Conditional probability is a measure of the probability of an event given that another event has already occurred.
a) True
b) False

Q73 Gaussian Naïve Bayes Classifier is _____distribution
a) Continuous
b) Discrete
c) Binary

Q74 Which of the following are several models for feature extraction
a) regression
b) classification
c) None of the above

Q75. The effectiveness of an SVM depends upon:
a) Selection of Kernel
b) Kernel Parameters
c) Soft Margin Parameter C
d) All of the above

Q76 RNNs stands for?

    a) Receives neural networks
    b) Report neural networks
    c) Recording neural networks
    d) Recurrent neural networks

Q77 CNN is mostly used when there is an?

    a) structured data
    b) unstructured data
    c) Both A and B
    d) None of the above

Q78 Which of the following is/are Limitations of deep learning?

    a) Data labelling
    b) Obtain huge training datasets
    c) Both A and B
    d) None of the above

Q79 What are order for the gradient descent algorithm?

1. Calculate error between the actual value and the predicted value
2. Reiterate until you find the best weights of network
3. Pass an input through the network and get values from output layer
4. Initialize random weight and bias
5. Go to each neuron which contributes to the error and change its respective values to reduce the error

    a) 1, 2, 3, 4, 5
    b) 5, 4, 3, 2, 1
    c) 3, 2, 1, 5, 4
    d) 4, 3, 1, 5, 2

Q 80. Which of the following techniques perform similar operations as dropout in a neural network?

    a) Bagging
    b) Boosting
    c) Stacking
    d) None of these