**Unit IV Naïve Bayes and Support Vector Machine**

1. How many terms are required for building a bayes model?

a) 1

b) 2

c) 3

d) 4

Answer: c

2. What is needed to make probabilistic systems feasible in the world?

a) Reliability

b) Crucial robustness

c) Feasibility

d) None of the mentioned

Answer: b

3. Where does the bayes rule can be used?

a) Solving queries

b) Increasing complexity

c) Decreasing complexity

d) Answering probabilistic query

Answer: d

4. What does the bayesian network provides?

a) Complete description of the domain

b) Partial description of the domain

c) Complete description of the problem

d) None of the mentioned

Answer: a

5. How the entries in the full joint probability distribution can be calculated?

a) Using variables

b) Using information

c) Both Using variables & information

d) None of the mentioned

Answer: b

6. How the bayesian network can be used to answer any query?

a) Full distribution

b) Joint distribution

c) Partial distribution

d) All of the mentioned

Answer: b

7. How the compactness of the bayesian network can be described?

a) Locally structured

b) Fully structured

c) Partial structure

d) All of the mentioned

Answer: a

8. To which does the local structure is associated?

a) Hybrid

b) Dependant

c) Linear

d) None of the mentioned

Answer: c

9. Which condition is used to influence a variable directly by all the others?

a) Partially connected

b) Fully connected

c) Local connected

d) None of the mentioned

Answer: b

10. What is the consequence between a node and its predecessors while creating bayesian network?

a) Functionally dependent

b) Dependant

c) Conditionally independent

d) Both Conditionally dependant & Dependant

Answer: c

11.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

Solution: B

12. When the C parameter is set to infinite, which of the following holds true?

A) The optimal hyperplane if exists, will be the one that completely separates the data

B) The soft-margin classifier will separate the data

C) None of the above

Solution: A

13.What do you mean by a hard margin?

A) The SVM allows very low error in classification

B) The SVM allows high amount of error in classification

C) None of the above

Solution: A

14.The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM’s?

A) Large datasets

B) Small datasets

C) Medium sized datasets

D) Size does not matter

Solution: A

15. The effectiveness of an SVM depends upon:

A) Selection of Kernel

B) Kernel Parameters

C) Soft Margin Parameter C

D) All of the above

Solution: D

16. Support vectors are the data points that lie closest to the decision surface.

A) TRUE

B) FALSE

Solution: A

17. The SVM’s are less effective when:

A) The data is linearly separable

B) The data is clean and ready to use

C) The data is noisy and contains overlapping points

Solution: C

18. 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 modeling

B) The model would consider only the points close to the hyperplane for modeling

C) The model would not be affected by distance of points from hyperplane for modeling

D) None of the above

Solution: B

19. The cost parameter in the SVM means:

A) The number of cross-validations to be made

B) The kernel to be used

C) The tradeoff between misclassification and simplicity of the model

D) None of the above

Solution: C

20.Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it’s hyper parameter. Based upon that give the answer for following question.

What would happen when you use very large value of C(C->infinity)?

A) We can still classify data correctly for given setting of hyper parameter C

B) We can not classify data correctly for given setting of hyper parameter C

C) Can’t Say

D) None of these

Solution: A

21. What would happen when you use very small C (C~0)?

A) Misclassification would happen

B) Data will be correctly classified

C) Can’t say

D) None of these

Solution: A

22. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?

A) Underfitting

B) Nothing, the model is perfect

C) Overfitting

Solution: C

23. Which of the following are real world applications of the SVM?

A) Text and Hypertext Categorization

B) Image Classification

C) Clustering of News Articles

D) All of the above

Solution: D

24.Which of the following option would you more likely to consider iterating SVM next time?

A) You want to increase your data points

B) You want to decrease your data points

C) You will try to calculate more variables

D) You will try to reduce the features

Solution: C

26. We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?

1. We do feature normalization so that new feature will dominate other

2. Some times, feature normalization is not feasible in case of categorical variables

3. Feature normalization always helps when we use Gaussian kernel in SVM

A) 1

B) 1 and 2

C) 1 and 3

D) 2 and 3

Solution: B

Question context: 27 – 28

Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

27) Now, think that you increase the complexity(or degree of polynomial of this kernel). What would you think will happen?

A) Increasing the complexity will overfit the data

B) Increasing the complexity will underfit the data

C) Nothing will happen since your model was already 100% accurate

D) None of these

Solution: A

28) In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?

1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data

2. Since data is fixed and SVM doesn’t need to search in big hypothesis space

A) 1

B) 2

C) 1 and 2

D) None of these

Solution: C

29. What is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space

2. It’s a similarity function

A) 1

B) 2

C) 1 and 2

D) None of these

Solution: C

**UNIT V**

**Decision Trees and Ensemble Learning**

1. Predicting with trees evaluate \_\_\_\_\_\_\_\_\_\_\_\_\_ within each group of data.

a) equality

b) homogeneity

c) heterogeneity

d) all of the mentioned

Answer: b

2. Point out the wrong statement.

a) Training and testing data must be processed in different way

b) Test transformation would mostly be imperfect

c) The first goal is statistical and second is data compression in PCA

d) All of the mentioned

Answer: a

3. Which of the following method options is provided by train function for bagging?

a) bagEarth

b) treebag

c) bagFDA

d) all of the mentioned

Answer: d

4. Which of the following is correct with respect to random forest?

a) Random forest are difficult to interpret but often very accurate

b) Random forest are easy to interpret but often very accurate

c) Random forest are difficult to interpret but very less accurate

d) None of the mentioned

Answer: a

5. Point out the correct statement.

a) Prediction with regression is easy to implement

b) Prediction with regression is easy to interpret

c) Prediction with regression performs well when linear model is correct

Answer: d

6. Which of the following library is used for boosting generalized additive models?

a) gamBoost

b) gbm

c) ada

d) all of the mentioned

Answer: a

7. The principal components are equal to left singular values if you first scale the variables.

a) True

b) False

Answer: b

8. Which of the following is statistical boosting based on additive logistic regression?

a) gamBoost

b) gbm

c) ada

d) mboost

Answer: a

9. Which of the following is one of the largest boost subclass in boosting?

a) variance boosting

b) gradient boosting

c) mean boosting

d) all of the mentioned

Answer: b

10. PCA is most useful for non linear type models.

a) True

b) False

Answer: b

11.varImp is a wrapper around the evimp function in the \_\_\_\_\_\_\_ package.

a) numpy

b) earth

c) plot

d) none of the mentioned

Answer: b

12. Point out the wrong statement.

a) The trapezoidal rule is used to compute the area under the ROC curve

b) For regression, the relationship between each predictor and the outcome is evaluated

c) An argument, para, is used to pick the model fitting technique

d) All of the mentioned

Answer: c

13. Which of the following curve analysis is conducted on each predictor for classification?

a) NOC

b) ROC

c) COC

d) All of the mentioned

Answer: b

14. Which of the following function tracks the changes in model statistics?

a) varImp

b) varImpTrack

c) findTrack

d) none of the mentioned

Answer: a

15. Point out the correct statement.

a) The difference between the class centroids and the overall centroid is used to measure the variable influence

b) The Bagged Trees output contains variable usage statistics

c) Boosted Trees uses different approach as a single tree

d) None of the mentioned

Answer: a

16. The advantage of using a model-based approach is that is more closely tied to the model performance.

a) True

b) False

Answer: a

17. Which of the following model sums the importance over each boosting iteration?

a) Boosted trees

b) Bagged trees

c) Partial least squares

d) None of the mentioned

Answer: a

18. Which of the following argument is used to set importance values?

a) scale

b) set

c) value

d) all of the mentioned

Answer: a

19. For most classification models, each predictor will have a separate variable importance for each class.

a) True

b) False

Answer: a

20. A \_\_\_\_\_\_\_\_\_ is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

a) Decision tree

b) Graphs

c) Trees

d) Neural Networks

Answer: a

21. Decision Tree is a display of an algorithm.

a) True

b) False

Answer: a

22. What is Decision Tree?

a) Flow-Chart

b) Structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label

c) Flow-Chart & Structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label

d) None of the mentioned

Answer: c

23. Decision Trees can be used for Classification Tasks.

a) True

b) False

View Answer

Answer: a

24. Choose from the following that are Decision Tree nodes?

a) Decision Nodes

b) End Nodes

c) Chance Nodes

d) All of the mentioned

Answer: d

25. Decision Nodes are represented by \_\_\_\_\_\_\_\_\_\_\_\_

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: b

26. Chance Nodes are represented by \_\_\_\_\_\_\_\_\_\_

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: c

27. End Nodes are represented by \_\_\_\_\_\_\_\_\_\_

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: d

28. Which of the following are the advantage/s of Decision Trees?

a) Possible Scenarios can be added

b) Use a white box model, If given result is provided by a model

c) Worst, best and expected values can be determined for different scenarios

d) All of the mentioned

Answer: d

29.Which of the following algorithm is not an example of an ensemble method?

A. Extra Tree Regressor

B. Random Forest

C. Gradient Boosting

D. Decision Tree

Solution: (D)

30. What is true about an ensembled classifier?

1. Classifiers that are more “sure” can vote with more conviction

2. Classifiers can be more “sure” about a particular part of the space

3. Most of the times, it performs better than a single classifier

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. All of the above

Solution: (D)

31. Which of the following option is / are correct regarding benefits of ensemble model?

1. Better performance

2. Generalized models

3. Better interpretability

A. 1 and 3

B. 2 and 3

C. 1 and 2

D. 1, 2 and 3

Solution: (C)

32.Which of the following can be true for selecting base learners for an ensemble?

1. Different learners can come from same algorithm with different hyper parameters

2. Different learners can come from different algorithms

3. Different learners can come from different training spaces

A. 1

B. 2

C. 1 and 3

D. 1, 2 and 3

Solution: (D)

33. True or False: Ensemble learning can only be applied to supervised learning methods.

A. True

B. False

Solution: (B)

34. True or False: Ensembles will yield bad results when there is significant diversity among the models.

A. True

B. False

Solution: (B)

35.Which of the following is / are true about weak learners used in ensemble model?

1. They have low variance and they don’t usually overfit

2. They have high bias, so they can not solve hard learning problems

3. They have high variance and they don’t usually overfit

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. None of these

Solution: (A)

36.True or False: Ensemble of classifiers may or may not be more accurate than any of its individual model.

A. True

B. False

Solution: (A)

37.parameters of all base models to improve the ensemble performance?

A. Yes

B. No

C. can’t say

Solution: (B)

38. Generally, an ensemble method works better, if the individual base models have \_\_\_\_\_\_\_\_\_\_\_\_?

A. Less correlation among predictions

B. High correlation among predictions

C. Correlation does not have any impact on ensemble output

D. None of the above

Solution: (A)

39. Which of the following ensemble method works similar to above-discussed election procedure?

A. Bagging

B. Boosting

C. A Or B

D. None of these

Solution: (A)

40. Suppose you are given ‘n’ predictions on test data by ‘n’ different models (M1, M2, …. Mn) respectively. Which of the following method(s) can be used to combine the predictions of these models?

1. Median

2. Product

3. Average

4. Weighted sum

5. Minimum and Maximum

6. Generalized mean rule

A. 1, 3 and 4

B. 1,3 and 6

C. 1,3, 4 and 6

D. All of above

Solution: (D)

41. If you want to ensemble these models using majority voting method. What will be the maximum accuracy you can get?

A. 100%

B. 78.38 %

C. 44%

D. 70

Solution: (A)

42. If you want to ensemble these models using majority voting. What will be the minimum accuracy you can get?

A. Always greater than 70%

B. Always greater than and equal to 70%

C. It can be less than 70%

D. None of these

Solution: (C)

43. How can we assign the weights to output of different models in an ensemble?

1. Use an algorithm to return the optimal weights

2. Choose the weights using cross validation

3. Give high weights to more accurate models

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. All of above

Solution: (D)

44. Which of the following is true about averaging ensemble?

A. It can only be used in classification problem

B. It can only be used in regression problem

C. It can be used in both classification as well as regression

D. None of these

Solution: (C)

**UNIT VI**

**Clustering Techniques**

1. Which of the following clustering type has characteristic shown in the below figure?  
  
a) Partitional  
b) Hierarchical  
c) Naive bayes  
d) None of the mentioned  
Answer: b

2. Point out the correct statement.  
a) The choice of an appropriate metric will influence the shape of the clusters  
b) Hierarchical clustering is also called HCA  
c) In general, the merges and splits are determined in a greedy manner  
d) All of the mentioned  
Answer: d

3. Which of the following is finally produced by Hierarchical Clustering?  
a) final estimate of cluster centroids  
b) tree showing how close things are to each other  
c) assignment of each point to clusters  
d) all of the mentioned  
Answer: b

4. Which of the following is required by K-means clustering?  
a) defined distance metric  
b) number of clusters  
c) initial guess as to cluster centroids  
d) all of the mentioned  
Answer: d

5. Point out the wrong statement.  
a) k-means clustering is a method of vector quantization  
b) k-means clustering aims to partition n observations into k clusters  
c) k-nearest neighbor is same as k-means  
d) none of the mentioned  
Answer: c

6. Hierarchical clustering should be primarily used for exploration.  
a) True  
b) False  
Answer: a

8. Which of the following function is used for k-means clustering?  
a) k-means  
b) k-mean  
c) heatmap  
d) none of the mentioned  
Answer: a

9. Which of the following clustering requires merging approach?  
a) Partitional  
b) Hierarchical  
c) Naive Bayes  
d) None of the mentioned  
Answer: b

10. K-means is not deterministic and it also consists of number of iterations.  
a) True  
b) False  
Answer: a