|  |  |  |  |
| --- | --- | --- | --- |
| **UNIT II** | | | |
| Comparison between regression and machine learning models- Compensating factors in machine learning models- Assumptions of linear regression Steps applied in linear regression modelling - Machine learning models - ridge and lasso regression - Example of ridge regression machine learning, Example of lasso regression machine learning mode - Logistic Regression Versus Random Forest-Maximum likelihood estimation - Terminology involved in logistic regression - Applying steps in logistic regression modelling - Random forest - Example of random forest using German credit data - Grid search on random forest - Variable importance plot - Comparison of logistic regression with random forest | | | |
| **PART-A (Multiple Choice Questions)** | | | |
| **Q.No.1** | **Questions** | **Course Outcome** | **Competence**  **BT Level** |
| **1** | Which of the following methods do we use to find the best fit line for data in Linear Regression?  **A) Least Square Error** B) Maximum Likelihood C) Logarithmic Loss D) Area Under Curve | 2 | 1 |
| **2** | Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?  A) AUC-ROC B) Accuracy C) Logloss **D) Mean-Squared-Error** | 2 | 1 |
| **3** | Suppose that we have N independent variables (X1,X2… Xn) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data.  You found that correlation coefficient for one of it’s variable(Say X1) with Y is -0.95.  Which of the following is true for X1?  A) Relation between the X1 and Y is weak **B) Relation between the X1 and Y is strong** C) Relation between the X1 and Y is neutral D) Correlation can’t judge the relationship | 2 | 3 |
| **4** | We can also compute the coefficient of linear regression with the help of an analytical method called “Normal Equation”. Which of the following is/are true about Normal Equation?  1. We don’t have to choose the learning rate  2. It becomes slow when number of features is very large  3. There is no need to iterate   A) 1 and 2 B) 1 and 3 C) 2 and 3 **D) 1,2 and 3** | 2 | 3 |
| **5** | What will happen when you apply very large penalty?  A) Some of the coefficient will become absolute zero **B) Some of the coefficient will approach zero but not absolute zero**  C) Linear regression is sensitive to outliers D) Linear regression is not sensitive to outliers | 2 | 2 |
| **6** | In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?  A) Bias will be high, variance will be high B) Bias will be low, variance will be high **C) Bias will be high, variance will be low** D) Bias will be low, variance will be low | 2 | 2 |
| 7 | Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?   A) Increase B) Decrease C) Remain constant **D) Can’t Say** | 2 | 3 |
| **8** | What do you expect will happen with bias and variance as you increase the size of training data?  A) Bias increases and Variance increases B) Bias decreases and Variance increases C) Bias decreases and Variance decreases **D) Bias increases and Variance decreases** | 2 | 3 |
| **9** | What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?  A)Less than 0 B) Greater than zero **C) Equal to 0** D) Greater than or equal to zero | 2 | 2 |
| **10** | Suppose you got the tuned hyper parameters from the previous question. Now, Imagine you want to add a variable in variable space such that this added feature is important. Which of the following thing would you observe in such case?  A) Training Error will decrease and Validation error will increase  B) Training Error will increase and Validation error will increase C) Training Error will increase and Validation error will decrease **D) Training Error will decrease and Validation error will decrease** | 2 | 1 |
| **11** | Regression models a target prediction value based on \_\_\_\_\_.  A) dependent variable **B) independent variables** C) independent value D) dependent value | 2 | 1 |
| **12** | Which of the following is a disadvantage of non-parametric machine learning algorithms?  a) Capable of fitting a large number of functional forms (flexibility)  b) Very fast to learn (speed)  **c) More of a risk to overfit the training data (overfitting)**  d) They do not require much training data | 2 | 1 |
| **13** | The scikit-learn Python machine learning library provides an implementation of the Ridge Regression algorithm via the Ridge class. Confusingly, the lambda term can be configured via the “\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_” argument when defining the class.  a) Lambda.  b) Gamma.  c) Beta.  **d) Alpha.** | 2 | 1 |
| **14** | Ridge regression can reduce the slope close to zero (but not exactly zero) but Lasso regression can reduce the slope to be exactly equal to zero.  **a) Both statements are True about Ridge and Lasso**.  b) Both statements are False about Ridge and Lasso.  c) True statement about Ridge but not about Lasso.  d) True statement about Lasso but not about Ridge. | 2 | 2 |
| 15 | The following statement isI. Lasso Regression helps to reduce overfitting and it is particularly useful for feature selection. II. Lasso regression can be useful if we have several independent variables that are useless.  * + a) Statement ( I ) is true and statement ( II ) is false.   + b) Statement ( I ) is false and statement ( II ) is true.   + c) Both Statement ( I ) & ( II ) are wrong.   + **D) Both Statement ( I ) & ( II ) are true.** | 2 | 1 |
| **16** | * + Which of the following is used where the target variable is of categorical nature? A. Keras B. Knime **C. Logistic Regression** D. MXNet | 2 | 1 |
| **17** | * + How many different types of Logistic Regression? A. 2 **B. 3** C. 4 D. 5 | 2 | 1 |
| **18** | * + \_\_\_\_\_\_\_ are defined as the ratio of the probability of an event occurring to the probability of the event not occurring. A. Simple B. Even C. Regex **D. Odds** | 2 | 1 |
| **19** | * + Logistic regression is used when you want to:   + **A. Predict a dichotomous variable from continuous or dichotomous variables.**   + B. Predict a continuous variable from dichotomous variables.   + C. Predict any categorical variable from several other categorical variables.   + D. Predict a continuous variable from dichotomous or continuous variables. | 2 | 2 |
| **20** | * + \_\_\_\_\_\_\_helps improve machine learning results by combining several models.   + A) Machine Learning B) bagging C) Entropy **D) Ensemble learning** | 2 | 1 |
| **21** | * + Which of the following is statistical boosting based on additive logistic regression? **a) gamBoost** b) gbm c) ada d) mboost | 2 | 1 |
| * + 22 | * + The measures developed for selecting the best split are often based on the degree of impurity of the child nodes. Which of the following is NOT an impurity measure?   + a) Gini   + b) Entropy   + **c) Pruning**   + d) Classification error | * + 2 | * + 1 |
| **23** | * + We are dealing with samples x where x is a single value. We would like to test two alternative regression models:   + 1) y = ax + e   + 2) y = ax + bx2 + e   + Which of the two models is more likely to fit the test data better?   + a) model 1   + b) model 2   + c) both will equally fit   + **d) impossible to decide** | 2 | 3 |
| **24** | * + In the general case, imagine that we have d binary features, and we want to count the number of features with value 1. How many leaf nodes would a decision tree need to represent this function?   + a) 21 leaf nodes   + **b) 2d leaf nodes**   + c) 2d-1leaf nodes   + d) 2d-1 leaf nodes | 2 | 2 |
| **25** | * + Which of the following classifiers can generate linear decision boundary?   + **a) Linear SVM**   + b) Random forest   + c) Naïve Bayes   + d) k-NN | 2 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| **UNIT III** | | | |
| K-nearest neighbors-KNN voter example - Curse of dimensionality-Curse of dimensionality with 1D, 2D, and 3D example. Curse of dimensionality with 3D example. KNN classifier with breast cancer Wisconsin data example- Naive Bayes - Probability fundamentals - Joint probability- Understanding Bayes theorem with conditional probability- Naive Bayes classification- Laplace estimator- Naive Bayes SMS spam classification example. | | | |
| **PART-A (Multiple Choice Questions)** | | | |
| **Q.**  **No** | **Questions** | **Course Outcome** | **Competence**  **BT Level** |
| **1** | * + Time to classify a new example than with a model in KNN requires?   + a. Depends on Data   + **b. More Time**   + c. None of these   + d. Less time | 3 | 2 |
| **2** | * + The Euclidean distance between two a set of numerical attributes is called as?   + **a. Closeness**   + b. Validation data   + c. Error Rate   + d. testing data | 3 | 1 |
| **3** | * + Which capture the local structure in data ?   + **a. Low Value**   + b. High Value   + c. Error Rate   + d. Median value | 3 | 1 |
| **4** | * + Which provide more smoothing, less noise ?   + a. Low Value   + **b. High Value**   + c. Error Rate   d. Median value | 3 | 1 |
| **5** | * + Which of the following option is true about k-NN algorithm?   + A) It can be used for classification B) It can be used for regression **C) It can be used in both classification and regression** | 3 | 1 |
| **6** | * + Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?   + **A) K-NN** B) Linear Regression C) Logistic Regression | 3 | 1 |
| **7** | * + Which of the following distance measure do we use in case of categorical variables in k-NN?   + 1.Hamming Distance   + 2. Euclidean Distance   + 3. Manhattan Distance   + **A) 1** B) 2 C) 3 D) 1 and 2 E) 2 and 3   + F) 1,2 and 3 | 3 | 1 |
| **8** | What is the naïve assumption in a Naïve Bayes Classifier   |  |  | | --- | --- | | A. | all the classes are independent of each other | | B. | all the features of a class are independent of each other | | C. | the most probable feature for a class is the most important feature to be cins3idered for classification | | **D.** | **all the features of a class are conditionally dependent on each other3333** | | 3 | 2 |
| **9** | Spam Classification is an example for ?  **a. Naive Bayes**  b. Probabilistic condition  c. Random Forest | 3 | 2 |
| **10** | Time complexity for Naive Bayes classifier for n feature, L classdata is  a. n\*L  b . O(n+L)  **c. O(n\*L)**   d. O(n/L) | 3 | 3 |
| **11** | A list of symptoms, predict whether a patient has diseaseX or not  **a. Medical Diagnosis**  b. Weather Diagnosis  c. Spam Diagnosis | 3 | 2 |
| **12** | How many Firms in Exact Bayes Calculation ?  a. One  **b. Two**  c. Three  d. Four | 3 | 1 |
| **13** | In Naive Bayes the relation between P(Fraud) and P(Truthful) is ?  Options :  **a. Greater than**  b. Lesser than  c. Equal to  d. Equal to 1 | 3 | 3 |
| **14** | If a bag contains red and blue balls and randomly picked 10 balls one by one wit replacement and out of 10, 3 red balls appear in trials, then the probability of red ball is  a. 0.6  **b. 0.3**  c. 0.4  d. 0.5 | 3 | 3 |
| **15** | If a trial has two outcomes then the events are considered as \_\_\_\_\_\_\_\_\_.  a. double events  b. success event  **c. mutually exclusive**  d. failure event | 3 | 2 |
| **16** | Naïve Bayes is \_\_\_\_\_\_\_\_\_\_  **a. Conditional Independence**  b. Conditional Dependence  c. Both a and b  d. None of the above | 3 | 1 |
| **17** | Spam Classification is an example for  a. Probabilistic condition  b. Random forest  **c. Naive bayes**  d. All the above | 3 | 2 |
| **18** | Which preprocessing technique in NLP used to reduce the dimensionality?  a. Word tokenization  **b. Lemmatization of words**  c. Part-of-Speech tagging  d. Removal of punctuation | 3 | 2 |
| **19** | Naïve Bayes requires \_\_\_\_\_\_\_\_\_\_  **a. Categorical values**  b. Numerical values  c. Either a or b  d. Both a and b | 3 | 1 |
| **20** | The probability of an event before evidence is seen is \_\_\_\_\_\_\_  a. Posteriori probability  **b. Prior probability**  c. Likelihood  d. conditional probability | 3 | 1 |
| **21** | **\_\_\_\_\_\_\_\_\_\_\_** is the likelihood of an outcome occurring based on the previous outcome occurring.  a. Posteriori probability  b. Prior probability  c. marginal probability  **d. conditional probability** | 3 | 1 |
| **22** | \_\_\_\_\_\_\_\_\_\_\_\_\_ is a statistical measure that evaluates how relevant a word is to a document in a collection of documents.  a. Term frequency  b. Inverse document frequency  **c. Term frequency-Inverse document frequency**  d. document frequency | 3 | 1 |
| **23** | Which function describes the joint probability of the observed data as a function of the parameters of the chosen statistical model?  **a. Likelihood**  b. Independence  c. Dependence  d. Marginal | 3 | 1 |
| **24** | Which package is used for all preprocessing stages in Natural Language  Processing?  a. stats  **b. nltk**  c. csv  d. nlp | 3 | 1 |
| **25** | Which technique is used to tackle the problem of zero probability in Naïve Bayes machine learning algorithm?  a. Simple exponential  b. Random Walk  c. Moving Average  **d. Laplace estimator** | 3 | 1 |