ML Assignment – 2

1. what is regression analysis?
   Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and on or more independent variables is called regression analysis.
2. Explain the difference between linear and nonlinear regression.
   Linear regression:
   - Linear regression assumes the linear relationship between the independent and dependent variables.
   - The model fits a straight line to the data.

   Non-linear regression:

   -the nonlinear regression models a non-linear relationship meaning the change in the dependent variable is not constant for the same change in the independent variables.

   -the data fits a curved line instead of a straight one.

3. What is the difference between simple linear regression and multiple linear regression?
   Simple linear regression:
   - one dependent variable y predicted from one independent variable x.
   - one regression coefficient.
   - r square: proportion of variable in dependent variable y prediction from x.

   Multiple linear regression:
   -one dependent variable y predicted from a set of independent variables(x1,x2,…..xk).
   -one regression coefficient for each independent variable.
   - R square: proportion of variation in dependent variable y predictable by a set of independent variables(x's).
4. How is the performance of the regression model Typically evaluated?
   The performance of a regression model is typically evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). These metrics help measure the accuracy of the model's predictions compared to actual values, with lower MAE and MSE indicating better performance, and R^2 indicating the proportion of variance explained by the model.
5. What is overfitting in the context of the regression model?
   Overfitting in the context of a regression model occurs when the model learns not only the underlying pattern in the training data but also the noise and random fluctuations. This leads to high accuracy on the training data but poor generalization to new, unseen data, resulting in decreased predictive performance.
6. What is logistic regression used for?
   Logistic regression is used for binary classification problems, where the goal is to predict one of two possible outcomes. It estimates the probability that a given input belongs to a particular class, using a logistic function to model the relationship between the input features and the probability of the target class.
7. How does logistic regression differ from linear regression?
   Logistic regression differs from linear regression in that it is used for classification tasks rather than predicting continuous outcomes. While linear regression outputs a continuous value, logistic regression outputs probabilities that are mapped to binary classes using a logistic (sigmoid) function.
8. Explain the concept of odds ratio in logistic regression.
   In logistic regression, the odds ratio represents the change in odds of the outcome occurring for a one-unit increase in the predictor variable. It quantifies how the likelihood of the dependent event changes with changes in the independent variable, with values greater than 1 indicating increased odds and values less than 1 indicating decreased odds.

9. What is the sigmoid function in logistic regression?
   The sigmoid function in logistic regression is a mathematical function that maps any real-valued number into a value between 0 and 1. It is defined as $\sigma(x)=1/1+e-x$ and is used to model the probability that a given input belongs to a particular class.
10. How is the performance of a logistic regression model evaluated?
    The performance of a logistic regression model is evaluated using metrics such as accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve (AUC-ROC). These metrics help assess how well the model distinguishes between the two classes, balancing the trade-offs between different types of errors.
11. What is a decision tree?
    A decision tree typically starts with a single node, which branches into possible outcomes. Ach of those outcomes leads to additional nodes, which branch off into other possibilities is called a decision tree.
12. What does a decision tree make predictions?
    A decision tree is a type of supervised machine learning used to categorical or make predictions based on how a previous set of questions were answered. This model is used for the training and tested on a set of data that contains the design categorization.
13. What is entropy in the context of a decision tree?
    the context of a decision tree, entropy is a measure of the impurity or randomness in a set of data. It quantifies the amount of uncertainty or disorder in the dataset, with higher entropy indicating more mixed or uncertain class distributions. Entropy is used to determine the best split points in the tree, aiming to reduce the impurity and create more homogeneous subsets.
14. What is pruning in a decision tree?
    Pruning is a technique that removes the parts of the decision tree which prevents it from growing to its full depth. The parts that it removes from the tree are the parts that do not provide the power to classify instances.
15. How do the decision tree handle missing values?
    To handle the missing values effectively, decision trees use surrogate splits. These splits act as a backup choice when the primary attribute for a split has missing values. The algorithm identifies the next best attribute that can provide a similar separation as the primary attribute.
16. What is a support vector machine (SVM)?
    A support vector machine is a supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an n-dimensional space.
17. Explain the concept of margin in SVM.
    Margin is the distance between the hyperplane and the observations closest to the hyperplane.in SVM large margin is considered a good margin. The main objective of the support vector machine algorithm is to maximize the margin.
18. What are support vectors in SVM?
    Support vectors in Support Vector Machines (SVM) are the data points that lie closest to the decision boundary (or hyperplane). These points are critical in defining the position and orientation of the hyperplane, as they directly influence the margin and are used to maximize the distance between the classes.
19. How does SVM handle non-linearly separable data?
    SVM handles non-linearly separable data by using the kernel trick. This involves transforming the input data into a higher-dimensional space where a linear separation is possible. Common kernels include polynomial, radial basis function (RBF), and sigmoid, which enable SVM to find an optimal hyperplane in the transformed feature space.
20. What are the advantages of SVM over other classification algorithms?
    The advantages of SVM over other classification algorithms.
    -effective in high-dimensional spaces.
    -robust against overfitting

-versatile with different kernels.
-effective in cases of small data sets.
-global solutions
-scalability

21. What is the naïve bayes algorithms?

It is the classification technique based on Baye's theorem with an independence assumption among predictors is called naïve bayes algorithm.

22. Why is it called "naive" Bayes?

Naïve Bayes is called naïve because it assumes that each input variable is independent. This is a strong assumption and unrealistic for real data, the technique is very effective on a large range of complex problems.

23. How does naïve Bayes handle continuous and categorical features?

Continuous Features: For continuous features, Naïve Bayes assumes a Gaussian (normal) distribution of the data within each class. It estimates the mean and variance of each feature for each class and then uses these parameters to compute the probability density function (PDF) of the feature value given the class.
Categorical Features: For categorical features, Naïve Bayes computes the frequency of each category within each class. It then calculates the conditional probability of a given feature value (category) given the class, using the frequency counts.

24. Explain the concept of prior and posterior probabilities in naïve Bayes.

Prior Probability: The prior probability P(C)represents the probability of class C occurring before considering any evidence or features. It reflects our initial belief about the likelihood of each class before observing the data.
Posterior Probability: The posterior probability P(C|X) represents the probability of class C given the observed features X. It is computed using Bayes' theorem: $P(C|X)=P(X|C)\cdot P(C)P(X)$

25. What is Laplace smoothing? How is it used in naïve Bayes?

Laplace smoothing is a smoothing strategy that assists tackle the issue of no likelihood in the naïve Bayes with machine learning calculation. It is a fun and fascinating algorithms with will help in utilizing higher alpha qualities.

26. Can Naïve Bayes be used for regression tasks?

Naïve Bayes is a supervised classification algorithm that is used primarily for dealing with binary and multi-class classification problems, though with some modifications, it can also be used for solving regression problems.

27. How do you handle missing value in naïve Bayes?

Naïve Bayes Imputation (NBI) is used to fill in missing values by replacing the attribute information according to the probability estimate. The NBI process divides the whole data into two sub-sets is the complete data and data containing missing data. Complete data is used for the imputation process at the lost value.

28. What are some common applications of naïve Bayes?

Naïve Bayes is commonly used in various applications such as:
-Text classification
-Recommendation Systems
-Medical diagnosis
-Fraud detection
-customer segmentation
-real-time prediction.

29. Explain the concept of feature independence assumption in naïve Bayes.

The features of the data are conditionally independent of each other given the class label. Continuec features are normally distributed, if a feature is continuous,then it is assumed to be normally distributed within each class.

30. How does naïve Bayes handle categorical features with a large number of categories?

Naïve Bayes handles categorical features with a large number of categories by computing the frequency of each category within each class. It uses these frequencies to estimate the conditional probability of a feature value (category) given the class, which helps in calculating posterior probabilities during classification. Despite the large number of categories, Naïve Bayes maintains efficiency due to its independence assumption and straightforward probability calculations.

31. What is the curse of dimensionality, and how does it affect machine learning algorithms?

The curse of dimensionality refers to the phenomenon where the performance of machine learning algorithms deteriorates as the number of features or dimensions in the data increases. because of the increased computational complexity, increased data sparsity, and overfitting.

32. Explain the bias-variance trade-off and its implications for machine learning models.

The bias-variances trade off descries the relationship between a models complexity the accuracy of its predictions and how well it can make predictions on previously unseen data that were not used to train the model.

-bias creates consistent errors in the machine learning model, which represents a simpler machine learning model that is not suitable for a specific requirement.

-variance creates errors that lead to incorrect predictions seeing trends or data points that do not exist.

33. What is cross-validation, and why is it used?

Cross-validation is a technique for evaluating machine learning models by training several machine learning models on subsets of the available input data and evaluating them on the complementing subset of the data.it is uses cross-validation to detect overfitting and fitting to generalize a pattern.

34. Explain the feature between parametric and non-parametric machine learning algorithms.

Parametric machine learning algorithms.

-algorithms that simplify the function to a known form are called parametric machine learning algorithms.

-the algorithm involves two steps such as:

1.selct a form for the function.

2.learn the coefficients for the function from the training data.

-these method is easy to understand and interpret the results.

-models are very fast to learn from data.

-they do not require as much training data and can work well even if the fit to the data is not perfect.

Non-parametric machine learning algorithm:

-algorithms that do not make strong assumptions about the form of the mapping function are called non-parametric machine learning algorithms.

-Flexibility

-power

-it can result in higher performance models for predictions.

-it requires a lot more training data to estimate the mapping function.

-It has a lot slower to train as they often have far more parameters to train.

35. What is feature scaling, and why is it important in machine learning?

Feature scaling is the process of normalizing the range of features in a dataset. Real-world datasetsoften contain features that vary in degrees of magnitude, range, and units. Therefore, for machine learning models to interpret these features on the same scale, we need to perform feature scaling.

36. what is regularization, and why is it used in machine learning?

Regularization is a set of methods for reducing overfitting in machine learning models. Typically regularization trades a marginal decrease in training accuracy for an increase in generalizability. Regularization encompasses a range of techniques to correct for overfitting in machine learning.

37. Explain the concept of ensemble learning and give an example.

Ensemble learning is a form of hybrid learning system in which multiple analytics are combined intelligently with the purpose of obtaining better results than a single analytics can provide. Random forest is an example for the ensemble machine learning.

38. What is the difference between bagging and boosting?

Bagging: bagging is a learning approach that aids in enhancing the performance, execution, and precision of machine learning algorithm is called bagging.

Boosting: Boosting is an approach that iteratively modifies the weight of observation based on the last classification.it is the easiest method of merging predictions that belong to the same type.

39. What is the difference between a generative model and a discriminative model?

Generative model: generative models can generate the new instances.

Discriminative model: Discriminative models discriminate between different kinds of data instances.

40. Explain the concept of batch gradient descent and stochastic gradient descent?

Batch gradient descent: batch gradient descent computes the gradient of the cost function with respect to the model parameters using the entire training dataset in each iteration.

Stochastic gradient descent: stochastic gradient descent computes the gradient using only a single training example or a small subset of examples in each iteration.

41. What is the k-nearest neighbors (KNN) algorithm, and how does it work?

The k-nearest neighbors (KNN) algorithm is a non-parametric and instance-based learning method used for classification and regression tasks. It works by calculating the distance (typically Euclidean distance) between the query instance and all other instances in the training data. The predicted outcome is based on the majority class (for classification) or the average value (for regression) of the k nearest instances to the query instance.

42. What are the disadvantages of the k-nearest neighbors algorithm?

Disadvantages:

-computational expense
-slow speed
-memory and storage issues for large datasets
-susceptibility to the curse of dimensionality.

43. Explain the concept of one-hot encoding and its use in machine learning.

One-hot encoding is a techniques in machine learning that turns categorical data like colors into numerical for machines to understand.it creates new binary columns and it indicating the presence of each possible value from the original data.

44. What is feature selection, and why is it important in machine learning?

Feature learning is the process is used to make the process more accurates. It also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones.

45. Explain the concept of cross-entropy loss and its use in classification tasks.

Cross-entropy loss measures the performance of a classification model whose output is a probability values between 0 and 1. It also increases the predicted probability diverges from the actual label.

46. What is the difference between batch learning and online learning?

Batch learning can be used to build a model on a large and reliable data set. while online learning is better suited to finetune the model on new and changing datasets.

47. Explain the concept of grid search and its use in hyperparameter tuning.

Grid search employs an exhaustive search strategy, systematically exploring various combinations of specified hyperparameters and their default values. This approach involves tuning parameters, such as learning rate, through a cross-validates model, which assesses performance across different parameter settings.

48. What are the advantages and disadvantages of a decision tree?

Advantages:

-interpretability,
-efficient handling of incomplete data,
-intrinsic variable selection

Disadvantages:

-overfitting
-sensitivity to small variations in the data.

49. What is the difference between L1 and L2 regularization?

The difference between L1 and L2 regularization:
The L1 regularization penalizes the sum of absolute values of the weights. It solution is a sparse, whereas L2 regularization penalizes the sum of squares of the weights.it solution is a non-sparse.

50. What are some common preprocessing techniques used in machine learning?

Some of the common preprocessing techniques used in machine learning such as
-data wrangling
-data transformation
-data reduction
-feature selection
-feature scaling

51. What is the difference between a parametric and non-parametric algorithm? Give an example of each.

Parametric algorithm:

A learning model that summarizes data with a set of fixed parameters. parametric machine alearning algorithms are which optimize the function to a known form. It deals with discrete valuesa.examples are: Logistic regression, linear discriminant analysis, naïve Bayes, and simple neutral networks.

Non-parametric algorithm:

Nonparametric algorithm machine learning algorithms are those which do not make specific assumptions about the type of the mapping function. it deals with continuous values. Examples are: k-nearest neighbors, decision tree, and support vector machines.

52. Explain the bias-variance trade-off and how it relates to model complexity.

The bias-variance trade-off refers to the dilemma in machine learning where increasing model complexity reduces bias (error from erroneous assumptions) but increases variance (sensitivity to noise). Balanced model complexity is crucial: too simple models underfit, while overly complex ones overfit, finding patterns in noise rather than true relationships.

53. What are the advantages and disadvantages of using ensemble methods like random forests?

Advantages:

-Predicts valid classification results with minimum training time.
-random forest is a popular classifier in data mining.
-high accuracy
-robustness to noise
-Estimating feature importance.
-handles missing data and outliers.

Disadvantages:

-random forest performs poorly on datasets with class noise.
-computational complexity.
-memory usage
-prediction time
-lack of interpretability.
-overfitting

54. Explain the difference between bagging and boosting.

Bagging: bagging is a learning approach that aids in enhancing the performance, execution, and precision of machine learning algorithms is called bagging.

Boosting: Boosting is an approach that iteratively modifies the weight of observation based on the last classification.it is the easiest method of merging predictions that belong to the same type.

55. What is the purpose of hyperparameter tuning in machine learning?

Hyperparameters directly control model structure, function, and performance. Hyperparameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning and choosing appropriate hyperparameter values is crucial for success.

56. What is the difference between regularization and feature selection?

Feature selection: feature selection is also known as feature subset selection, variable selection, or attribute selection. This approach removes the dimensions from the input data and results in a reduced data set for model inference.

Regularization: regularization is the process of where we are constraining the solution space while doing optimization.

57. How does the Lasso(L1) regularization differ from ridge(L2) regularization?

Lasso (L1)regularization:

-lasso regression uses L1 regularization.
-it uses the absolute value of the coefficient.
-lasso regression can force them to be exactly zero.
-Lasso regression performs both regularization and feature selection, making it more suitable for high-dimensional datasets.
-lasso regression can lead to a sparse model, which means it can create a model with fewer features,

Ridge (L2) regularization:

-Ridge regression uses L2 regularization.
-it uses the square of coefficients.
-Ridge regression shrinks coefficients of less significant features towards zero.
-Ridge regression generally works better in scenarios where there are fewer significant features.
-Ridge regression does not produce sparse models.

58. Explain the concept of cross-validation and why is it used.

Cross-validation is a technique for evaluating machine learning models by training several machine learning models on subsets of the available input data and evaluating them on the complementing subset of the data.it is uses cross-validation to detect overfitting and fitting to generalize a pattern

59. What are some common evaluation metrics used for regression tasks?

Some common evaluation metrics used for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (Coefficient of Determination), and Mean Absolute Percentage Error (MAPE).

60. How does the k-nearest neighbors(KNN) algorithm make predictions?

It works by finding the K nearest points in the training dataset and uses their class to predict the class or value of a new data point. It can handle complex data and is also easy to implement, which is why KNN has become a popular tool in the field of artificial intelligence. this means that the new point is assigned a value based on how closely it resembles the points in the training set.

61. What is the feature scaling? and why is it important in machine learning.

Feature scaling is the process of normalizing the range of features in a dataset. Real-world datasets often contain features that vary in degrees of magnitude, range, and units. Therefore, for machine learning models to interpret these features on the same scale, we need to perform feature scaling.

62. What is the curse of dimensionality, and how does it affect machine learning algorithms?

The curse of dimensionality refers to the phenomenon where the performance of machine learning algorithms deteriorates as the number of features or dimensions in the data increases. because of the increased computational complexity, increased data sparsity, and overfitting.

63. How do the naïve Bayes algorithms handle categorical features?

Naïve Bayes algorithms handle categorical features by calculating the frequency of each category within each class from the training data. this type of naïve is mainly used in a binary categorical tagete column where the problem statement is no predict only yes or no.

64. Explain the concept of prior and posterior probabilities in naive Bayes.

Prior Probability: The prior probability P(C)represents the probability of class C occurring before considering any evidence or features. It reflects our initial belief about the likelihood of each class before observing the data.

Posterior Probability: The posterior probability P(C|X) represents the probability of class C given the observed features X. It is computed using Bayes' theorem: $P(C|X)=P(X|C)\cdot P(C)P(X)$

65. What is Laplace smoothing, and why is it used in naïve Bayes?

Laplace smoothing is a smoothing strategy that assists tackle the issue of no likelihood in the naïve Bayes with machine learning calculation. It is a fun and fascinating algorithms with will help in utilizing higher alpha qualities.

66. Can naïve Bayes handle continuous features?

Yes, the naïve bayes classifier for functional data can be used on continuous data. It can be used on continuous data by estimating class conditional probability using a probability distribution function.

67. What are the assumptions of the naïve Bayes algorithm?

The assumptions of the naïve bayes algorithm include:

- Feature independence
- Continuous features are normally distributed.
- Discrete features have multinomial distributions.
- features are equally important
- no missing data

68. How does naïve Bayes handle missing values?

Naïve Bayes Imputation (NBI) is used to fill in missing values by replacing the attribute information according to the probability estimate. The NBI process divides the whole data into two sub-sets is the complete data and data containing missing data. Complete data is used for the imputation process at the lost value.

69. What are some common applications of naïve Bayes?

Naïve Bayes is commonly used in various applications such as:

-Text classification
-Recommendation Systems
-Medical diagnosis
-Fraud detection
-customer segmentation
-real-time prediction.

70. Explain the difference between generative and discriminative models

Generative model: generative models can generate the new instances.

Discriminative model: Discriminative models discriminate between different kinds of data instances.

71. How does the decision boundary of a naïve Bayes classifier look like for binary classification tasks?

The decision boundary of a Naïve Bayes classifier for binary classification tasks is typically linear or piecewise linear. This boundary is determined by the likelihoods of the features for each class and the class priors. Since Naïve Bayes assumes feature independence, the decision boundary can be straight or consist of segments that reflect the contribution of individual features to the classification decision.

72. What are the differences between multinomial naïve Bayes and Gaussian naïve bayes.

Multinomial Naïve Bayes:

- Designed for discrete data, particularly suited for text classification where features represent the frequency of words or events.
- it Uses counts or frequencies of features within each class.

- it Calculates the probability of observing a given set of counts (features) given the class.

Gaussian Naïve Bayes:

- Designed for continuous data, assuming that the features follow a normal (Gaussian) distribution.
- it Uses the mean and variance of the features to calculate the probability densities for each class.
- it Calculates the probability of observing a continuous value given the class, using the Gaussian (normal) distribution formula.

73. How does naïve Bayes handle numerical instability issues?

The naive Bayes algorithm is a powerful and efficient probabilistic classifier. it handles numerical data by assuming distributions or applying transformations. Numerical stability is addressed by using logarithmic probabilities and Laplace smoothing is used to handle zero probabilities.

74. What is the Laplace correlation, and when is it used in naïve Bayes?

.Laplace smoothing is a smoothing technique that helps tackle the problem of zero probability in the naïve Bayes machine learning algorithm. when a categorical feature value does not appear in the training data for a given class. It works by adding a small constant to all feature counts, ensuring that no probability is ever zero. This helps to avoid issues during probability calculation and improves the robustness of the model, particularly in cases with small or sparse datasets.

75. Can naïve Bayes be used for regression tasks?

Naïve Bayes is a supervised classification algorithm that is used primarily for dealing with binary and multi-class classification problems, though with some modifications, it can also be used for solving regression problems.

76. Explain the concept of conditional independence assumptions in naïve Bayes.

The concept of conditional independence assumptions in Naïve Bayes refers to the assumption that all features are independent of each other given the class label. This means that the presence or value of a feature is assumed to be unrelated to the presence or value of any other feature, provided the class is known. This simplification allows the Naïve Bayes classifier to calculate the joint probability of the features by multiplying the individual conditional probabilities, making the computation efficient and straightforward.

77. How does naïve Bayes handle categorical features with a large number of categories?

Naïve Bayes handles categorical features with a large number of categories by computing the frequency of each category within each class. It uses these frequencies to estimate the conditional probability of a feature value (category) given the class, which helps in calculating posterior probabilities during classification. Despite the large number of categories, Naïve Bayes maintains efficiency due to its independence assumption and straightforward probability calculations.

78. What are some drawbacks of the naïve Bayes algorithms?

Some drawbacks of the naive Bayes algorithms include:

-the naïve Bayes algorithm has trouble with the ‘zero’ frequency problem.

-it will assume that all the attributes are independent, which rarely happens in real life.

-it will estimate things wrong sometimes, so you shouldn’t take its probability outputs seriously.

79. Explain the concept of smoothing in naïve bayes.

Laplace smoothing, is a technique used to handle zero probability issues in the estimation of conditional probabilities for categorical features. Without smoothing, if a feature value does not appear in the training data for a particular class, the estimated probability for that feature value given the class would be zero, potentially leading to problems in classification.

80. How does naïve Bayes handle imbalanced datasets?

The naïve Bayes algorithm is sensitive to imbalanced data because it relies on probabilities to make predictions. if the classes in the dataset are not represented equally, the probabilities calculated by the algorithm may be skewed, and as a result, the predictions made by the algorithm may not be accurate.