Assignment-2: Introduction to Machine Learning Techniques

1. Explore the diversity of data types utilized in machine learning and the functions performed by different algorithms.
2. Analyze the RIPPER algorithm in the context of rule-based classification. Discuss its strengths and limitations compared to other rule-based algorithms and its practical applications.
3. Compare the fundamental structure and splitting criteria used in regression trees and model trees. When would a model tree be more appropriate than a regression tree?
4. Compare and contrast Adaboost and Gradient Boosting methods. Explain how AdaBoost enhances the accuracy of decision trees in comparison to other boosting algorithms.
5. Describe how AdaBoost helps in improving the overall performance of machine learning models. Provide examples or scenarios where using AdaBoost can significantly enhance model accuracy.
6. Detail the process of adding regression capabilities to decision trees. What advantages does this addition offer in comparison to traditional decision trees?
7. Discuss the 1R algorithm's approach to generating rules for classification. Explain how it selects a single attribute for decision-making and its implications for model performance.
8. Discuss the main types of Gradient Boosting (e.g., Gradient Boosting Machines, XGBoost, LightGBM). Highlight the distinguishing features of each type.
9. Explore Gradient Boosting and its different types, elucidating their impact on improving the performance of weak learners.
10. Explain the Separate & Conquer approach and how it's employed in generating classification rules. Discuss the significance of 0R, 1R, and RIPPER algorithms in this context.
11. Explain the concept of AdaBoost and how it contributes to improving the accuracy of weak learners. Discuss its importance in the context of ensemble learning.
12. Explain the core differences between Multiple Linear Regression and Logistic Regression. When would you choose one over the other for a given problem?
13. Detail the fundamental principles of Multiple Linear Regression and Logistic Regression, highlighting their distinctions in application and underlying concepts.
14. Explain the differences in the handling of weak learners between Adaptive Boosting (AdaBoost) and Gradient Boosting.
15. Explain the basic concepts of Bayesian learning and how it forms the foundation for naive Bayes classification.
16. How does naive Bayes handle numeric features in classification tasks? Explain the methods or transformations that enable its use with numerical data.
17. How does regularization (e.g., Ridge or Lasso) impact Multiple Linear Regression and Logistic Regression models?
18. How does the incorporation of linear models into the branches of a model tree improve predictive power over traditional regression trees?
19. How does the Laplace smoothing technique address the zero-frequency problem in naive Bayes? Explain its significance in classification.
20. Discuss the naive Bayes algorithm and its classification process, highlighting its assumption of feature independence.

21. What are the key differences between Gaussian Naive Bayes and Multinomial Naive Bayes? When would you choose one over the other?
22. Describe the naive Bayes algorithm for classification, emphasizing conditional probability and the "naive" assumption, supported by an illustrative example.
23. Discuss the use of odds ratios in Logistic Regression and how they differ from coefficients in Multiple Linear Regression.
24. Describe the underlying assumptions of Multiple Linear Regression. How do these assumptions differ from those in Logistic Regression?
25. What is Latent Dirichlet Allocation (LDA) and how does it contribute to topic modeling? Explain its core principle.
26. Illustrate the generative process of LDA. How does it assign topics to documents and words to topics?
27. Explain the generative process of LDA for modeling topics in document collections, detailing how LDA assigns topics to words and documents.
28. Explain the process of fitting an LDA model to a given text dataset. What are the key parameters involved in this process?
29. How is a word cloud generated and what insights can be drawn from it in the context of topic modeling?
30. What defines deep learning and how does it differ from traditional machine learning approaches?
31. What's the difference between predicting things and using machines to learn? Can you give examples of both?
32. Explain how lazy learning, like the K-Nearest Neighbors (KNN) way, works.
33. How does boosting help trees get better at guessing things?
34. How do decision trees decide stuff, and what's this divide and conquer thing they do?
35. How are model trees and regular trees different? Tell me when model trees are super useful in real life.
36. Talk about weak learners, leftovers, speed of learning, and how they all fit into making boosting better. Also, list different kinds of boosting.
37. How do pre-pruning and post-pruning techniques enhance the quality of regression and model trees in decision tree construction?
38. Explain OneR and RIPPER classification rules and their applications in machine learning.
39. Explore the criteria for splitting used in regression trees and their influence on tree building and model accuracy.
40. Explain the steps involved in preparing data for machine learning, encompassing data structures and manipulation methods.
41. What characterizes KNN as a "lazy" learning algorithm, and what implications stem from this?
42. Contrast pre-pruning and post-pruning techniques in decision trees and explain their role in model construction.
43. Briefly highlight on how AdaBoost enhances model performance and its implications.
44. Identify the suitable machine learning technique for detecting risky bank loans and describe its application.