PW SKILLS

Assignment - 4

ML

1. **What are ensemble techniques in machine learning?**
2. **Explain bagging and how it works in ensemble techniques.**
3. **What is the purpose of bootstrapping in bagging?**
4. **Describe the random forest algorithm.**
5. **How does randomization reduce overfitting in random forests?**
6. **Explain the concept of feature bagging in random forests.**
7. **What is the role of decision trees in gradient boosting?**
8. **Differentiate between bagging and boosting.**
9. **What is the AdaBoost algorithm, and how does it work?**
10. **Explain the concept of weak learners in boosting algorithms.**
11. **Describe the process of adaptive boosting.**
12. **How does AdaBoost adjust weights for misclassified data points?**
13. **Discuss the XGBoost algorithm and its advantages over traditional gradient boosting.**
14. **Explain the concept of regularization in XGBoost.**
15. **What are the different types of ensemble techniques?**
16. **Compare and contrast bagging and boosting.**
17. **Discuss the concept of ensemble diversity.**
18. **How do ensemble techniques improve predictive performance?**
19. **Explain the concept of ensemble variance and bias.**
20. **Discuss the trade-off between bias and variance in ensemble learning.**
21. **What are some common applications of ensemble techniques?**
22. **How does ensemble learning contribute to model interpretability?**
23. **Describe the process of stacking in ensemble learning.**
24. **Discuss the role of meta-learners in stacking.**
25. **What are some challenges associated with ensemble techniques?**
26. **What is boosting, and how does it differ from bagging?**
27. **Explain the intuition behind boosting.**
28. **Describe the concept of sequential training in boosting.**
29. **How does boosting handle misclassified data points?**
30. **Discuss the role of weights in boosting algorithms.**
31. **What is the difference between boosting and AdaBoost?**
32. **How does AdaBoost adjust weights for misclassified samples?**

33. Explain the concept of weak learners in boosting algorithms.
34. Discuss the process of gradient boosting.
35. What is the purpose of gradient descent in gradient boosting?
36. Describe the role of learning rate in gradient boosting.
37. How does gradient boosting handle overfitting?
38. Discuss the differences between gradient boosting and XGBoost.
39. Explain the concept of regularized boosting.
40. What are the advantages of using XGBoost over traditional gradient boosting?
41. Describe the process of early stopping in boosting algorithms.
42. How does early stopping prevent overfitting in boosting?
43. Discuss the role of hyperparameters in boosting algorithms.
44. What are some common challenges associated with boosting?
45. Explain the concept of boosting convergence.
46. How does boosting improve the performance of weak learners?
47. Discuss the impact of data imbalance on boosting algorithms.
48. What are some real-world applications of boosting?
49. Describe the process of ensemble selection in boosting.
50. How does boosting contribute to model interpretability?
51. Explain the curse of dimensionality and its impact on KNN.
52. What are the applications of KNN in real-world scenarios?
53. Discuss the concept of weighted KNN.
54. How do you handle missing values in KNN?
55. Explain the difference between lazy learning and eager learning algorithms, and where does KNN fit in?
56. What are some methods to improve the performance of KNN?
57. Can KNN be used for regression tasks? If yes, how?
58. Describe the boundary decision made by the KNN algorithm.
59. How do you choose the optimal value of K in KNN?
60. Discuss the trade-offs between using a small and large value of K in KNN.
61. Explain the process of feature scaling in the context of KNN.
62. Compare and contrast KNN with other classification algorithms like SVM and Decision Trees.

63. How does the choice of distance metric affect the performance of KNN?
64. What are some techniques to deal with imbalanced datasets in KNN?
65. Explain the concept of cross-validation in the context of tuning KNN parameters.
66. What is the difference between uniform and distance-weighted voting in KNN?
67. Discuss the computational complexity of KNN.
68. How does the choice of distance metric impact the sensitivity of KNN to outliers?
69. Explain the process of selecting an appropriate value for K using the elbow method.
70. Can KNN be used for text classification tasks? If yes, how?
71. How do you decide the number of principal components to retain in PCA?
72. Explain the reconstruction error in the context of PCA.
73. What are the applications of PCA in real-world scenarios?
74. Discuss the limitations of PCA.
75. What is Singular Value Decomposition (SVD), and how is it related to PCA?
76. Explain the concept of latent semantic analysis (LSA) and its application in natural language processing.
77. What are some alternatives to PCA for dimensionality reduction?
78. Describe t-distributed Stochastic Neighbor Embedding (t-SNE) and its advantages over PCA.
79. How does t-SNE preserve local structure compared to PCA?
80. Discuss the limitations of t-SNE.
81. What is the difference between PCA and Independent Component Analysis (ICA)?
82. Explain the concept of manifold learning and its significance in dimensionality reduction.
83. What are autoencoders, and how are they used for dimensionality reduction?
84. Discuss the challenges of using nonlinear dimensionality reduction techniques.
85. How does the choice of distance metric impact the performance of dimensionality reduction techniques?
86. What are some techniques to visualize high-dimensional data after dimensionality reduction?
87. Explain the concept of feature hashing and its role in dimensionality reduction.
88. What is the difference between global and local feature extraction methods?
89. How does feature sparsity affect the performance of dimensionality reduction techniques?
90. Discuss the impact of outliers on dimensionality reduction algorithms.