

Assignment 3

1. Collinearity
2. Random Forest
3. Decision Tree are prone to overfit
4. Training data
5. Anamoly detection
6. Case based
7. Both a and b
8. Both a and b
9. 3
10. PCA
11. Neither feature nor number of groups is known
12. SVG
13. Underfitting
14. Supervised learning
15. Mean squared error
16. Nonlinear, binary
17. supervised learning
18. euclidean distance
19. removing columns which have too many missing values
20. hidden attribute.
21. SVM is highly flexible
22. Only 2
23. $\frac{4}{10} \log(\frac{6}{10}) + \frac{6}{10} \log(\frac{4}{10})$
24. weights are regularized with the l1 norm
25. Logistic regression and Gaussian discriminant analysis
26. Either 2 or 3
27. increase by 125 pound
28. Minimize the squared distance from the points
29. The attributes are not linearly related
30. Convolutional Neural Network