

### **SOLUTIONS FILE 3**

1) D

2) B

3) C

4) C

5) C

6) C

7) D

8) C

9) A

10) A

11) C

12) B

13) B

14) A

15) B

16) C

17) A

18) C

19) It depends on the characteristics of the data set and the specific goals of the analysis. Generally, all of these techniques can be useful for reducing the dimensions of a data set, but they may be more or less effective depending on the particular context.

A. Removing columns with too many missing values can be a useful technique for reducing dimensions, especially if the missing values are random or the missingness is not related to the outcome of interest. However, if the missing values are informative or systematic, removing these columns could result in biased or incomplete analysis.

B. Removing columns with high variance can also be useful for reducing dimensions, especially if the variance is due to noise or measurement error rather than meaningful variation in the data. However, if the high variance is

due to important information or features in the data, removing these columns could result in loss of information and potentially biased analysis.

C. Removing columns with dissimilar data trends can be useful for reducing dimensions, especially if the trends are unrelated to the outcome of interest or if the trends are redundant with other columns in the data set. However, if the trends are important for understanding the relationships in the data or if they contain valuable information, removing these columns could result in loss of information and potentially biased analysis.

Therefore, it is important to carefully consider the characteristics of the data set and the specific goals of the analysis before deciding which technique to use for dimension reduction. In some cases, a combination of these techniques or other techniques may be most effective.

20) C

21) A

22) D

23) A

24) A

25) If the newly added labeled data point is correctly classified and lies far away from the decision boundary, then it is unlikely to change the learnt decision boundary in any of the models mentioned in the options (A), (B), (C), or (D). However, the situation may be different if the newly added data point is incorrectly classified or lies near the decision boundary.

26) D

27) B

28) D

29) B

30) B