Machine Learning-Worksheet8

1. B
2. A
3. A
4. C
5. D
6. D
7. C
8. D,A
9. C,D
10. A,B
11. One –hot coding becomes messier when we have large data in terms of more independent variables because it will create many dummy columns which might become a concern in terms of dimensionality. One could probably opt label encoding .

12. When we have imbalanced data we can opt

- RandomOverSampling: Creates multiple data points as the actual one.
- ADASYN(Adaptive Synthetic): In Adasyn the data points are synthetically created not just copying the minority class.
- SMOTE(Synthetic Minority Oversampling Technique**:**
  It aims to balance class distribution by randomly increasing minority class examples by replicating them. SMOTE synthesises new minority instances between existing minority instances.

13. The key difference between ADASYN and SMOTE is that the former uses a density distribution, as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed distributions. The latter generates the same number of synthetic samples for each original minority sample.

14. GridSearchCV tries all the combinations of the values passed in the dictionary and evaluates the model for each combination using the Cross-Validation method. Hence after using this function we get accuracy/loss for every combination of hyperparameters and we can

choose the one with the best performance.Not very suitable for larger datasets because it needs very large memory.

15. a)RMSE
    b)R squared error
    c)MSE