MACHINE LEARNING

Answers

1. R residual is better, because, In general, the higher the R-squared, the better the model fits your data.

2. The overall sum of squares (TSS) measures how a great deal version there's in the located records, even as the residual sum of squares measures the version in the mistakess among the located records and modeled values.

3. To minimize the adjusted loss function and prevent underfitting.

4. Gini Impurity is a dimension used to construct Decision Trees to decide how the capabilities of a dataset need to cut up nodes to shape the tree.

5. Yes. Overfitting can be one problem that describes if your model no longer generalizes well. Overfitting happens when any learning processing overly optimizes

6. Ensemble strategies are strategies that create a couple of fashions after which integrate them to supply advanced results.

7. Bagging is a method for lowering prediction variance via way of means of generating extra information for education from a dataset via way of means of combining repetitions with combos to create multi-units of the authentic information.

Boosting is an iterative **approach** for adjusting an observation's weight primarily based totally at the preceding classification.

8. The out-of-bag (OOB) **mistakess** is the **common mistakess** for **every** calculated **the use of** predictions from the **timber that don't include of their** respective bootstrap sample. This **lets in** the RandomForestClassifier to be **match** and **proven while** being trained

9. That k-fold **move** validation is a **system** used to estimate the **ability** of the **version** on new data.

10. Hyperparameter tuning takes **gain** of the processing infrastructure of Google Cloud **to check exclusive** hyperparameter configurations **while education** your model. It can **provide you with** optimized values for hyperparameters, which maximizes your model's predictive accuracy.

11. A **gaining knowledge of charge this is** too **big** can **reason** the version to converge too **quick** to a suboptimal solution, **while a gaining knowledge of charge this is** too small can **reason** the **procedure** to get stuck.

12. Logistic regression is neither linear **neither is** it a classifier. The **concept of** a "selection boundary" has little to do with logistic regression, **that is as a** substitute an immediate opportunity estimation **technique** that separates predictions from **selection**.

13. AdaBoost is the primary designed boosting set of rules with a selected loss function. On the opposite hand, Gradient Boosting is a widespread set of rules that assists in looking the approximate answers to the additive modelling problem.

14. In information and gadget learning, the prejudice−variance tradeoff is the belongings of a version that the variance of the parameter expected throughout samples may be decreased through growing the prejudice in the expected parameters.

STATISTICS WORKSHEET 5

Answers

1. D

2. C

3. C

4. B

5. D

6. B

7. A

8. A

9. B

10. A