ARTIFICIAL INTELLIGENCE QUESTION BANK

Module-3

1. **If you are a data scientist asked to plan budget for advertising for the advertising example used in the text, list out the questions that you will analyze the data for to plan the advertising budget for different media.**
2. **Explain simple linear regression. How are the co-efficients estimated?**
3. **Elaborate how closeness is measured using least squares criterion**
4. **Explain with suitable mathematical expressions for Standard Error (SE), accuracy of estimating the linear regression coefficients β_0 , β_1**
5. **Describe step by step, how you would establish that a predictor variable has a linear relationship with the output variable in linear regression model. [Table 3.1]**

	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

6. **Explain using appropriate measures as to how you will assess the accuracy of a linear regression model.**
7. **Explain multiple linear regression model and the associated RSS metric for gauging the fit.**
8. **Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, andnewspaper, rather than in terms of the coefficients of the linear model. [Chapter-3 Ex-1]**

	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

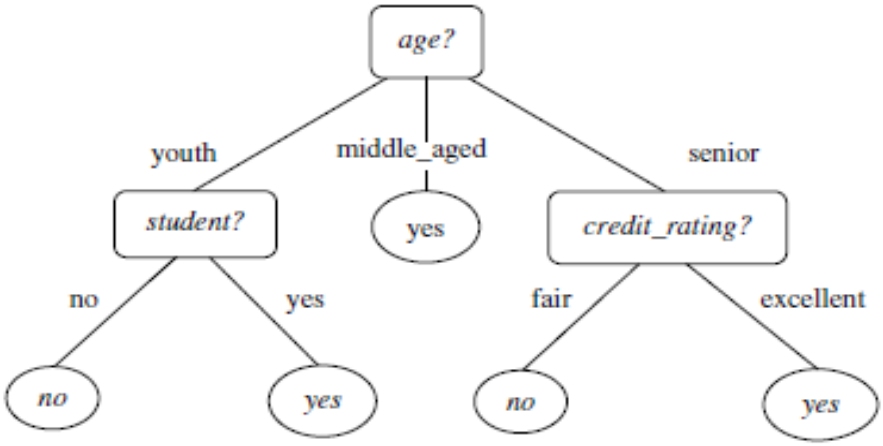
9. Describe the general approach to classification in machine learning.
10. Write the steps in decision tree construction.
11. How are decision trees used for classification? Explain with an example
12. Explain a. Gain Ratio

b. Gini Index

13. Explain common approaches of tree pruning
14. Why is tree pruning useful in decision tree induction? What is a drawback of using a separate set of tuples to evaluate pruning?
15. Why naïve Bayesian classification called “naïve”? Briefly outline the major steps of naïve Bayesian classification algorithm
16. Explain the different metrics for Evaluating Classifier performance.
17. Explain how well classifier can recognize tuples of different classes using confusion matrix
18. Explain the four additional aspects beyond accuracy-based measures that are used for Comparing classifiers.
19. Explain a. Handout method and Random Subsampling

    b. Cross validation

20. Give statistical tests of significance for model selection between any two classification models M1 and M2.
21. Compare classifiers based on cost-benefit and ROC curves
22. Construct a set of IF-THEN rules for the given decision tree [figure]



23. Compute the information gain and expected information of attribute age, needed to classify a tuple in *D* using the following Table

Class-Labeled Training Tuples from the *AllElectronics* Customer Database

RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

1.