

STATISTICS WORKSHEET 7

Ans 1(B)

Total number of times die is thrown =1402

Number of times outcomes is 6=190

Probability of the outcome 6 = number of times outcome

is / total number of times

die is thrown

$$= 190 / 1402 = 0.1355$$

Ans 2) (D)

$$52 + 44 + 20 + 56 + 40 = 212$$

Probability of getting a digit with unit place digit odd
number 1,3,5,7,9 = $212 / 400 = 0.53$

Ans 3) (C)

The probability that the tyre will last more than 9000
miles is .74545

$$375 + 445 / 1100$$

Ans 4)(B)

$$260+375/1100=0.577$$

The probability that if we buy a new tyre that it will last in the interval [4000-14000]miles is 0.577

Ans 5)(A)

$$10-4=5/10=0.5$$

Ans 6)(A)

Ans 7)(C)

$$\text{ie } 2/3=.33$$

Ans 8)(B)

Ans 9)(A)

Ans 10)(A)

Ans 11)(C)

Ans 12)(A)

Ans 13)(D)

Ans 14)(D)

Ans 15)(C)

MACHINE LEARNING

Ans 1)(A)

Ans 2)(A)

Ans 3)(B)

Ans 4)(A)

Ans 5)(A)

Ans 6)(B)

Ans 7)(B)

Ans 8)(A)

Ans 9)entropy= $\sum -P \cdot \log_2(P)$

Where p is the probability of entropy

Gini Index= $1 - \sum(P^2)$

Where p is the probability of object.

Ans 10)1)Random forests deals with the problem of overfitting as they combines the output of multiple decision trees.

2)while building a model using decision tree ,small changes in the data will bring a huge amount of

difference in predicting a model while this type of problem doesn't occur with RandomForest as data is sampled many times before model prediction.

Ans 11)It is the process of normalizing the range of features in the dataset.Feature scaling makes the training faster.Feature scaling is required to rescale the features.Scaling is the important step as it is used to treat the skewed data and rescale the features for modelling.

Techniques used for scaling are Standard Scaler and Min_Max scaler.

Ans 12)Gradient Descent is the most common optimization algorithm in machine learning .

- It is important to scale our inputs in Gradient Descent because scaling inputs makes gradient descent converge faster.
- Scaling the features in a machine learning model can improve the optimization process by making the flow of gradient descent smoother and helping

algorithm reach the minimum cost function more quickly.

- Feature scaling aims to speed up the process of convergence of Gradient Descent. It is an idea that makes sure that features that are involved in Gradient Descent are on the similar scale.

Ans 13) Since the accuracy model is easy to understand and use. It is most used metrics in machine learning .

But there are problems with it ,it cant be used for imbalanced dataset. In Machine learning classification accuracy just isn't the best option to check an imbalanced data .As the model will ignore smaller data and focus on the larger data to acquire higher accuracy .At the time of the conclusion ,this metric will ignore the small dataset .To avoid this confusion it is better to use F1 score .

Ans 14) Fscore metric is a measure to know the accuracy of the model. Also it conveys the balance

between precision and recall .0.0 is considered as the worst f1_score whereas 1.0 is a perfect f_1 score.F1 score is useful if you have uneven class distribution.

Mathematical formulae:

$$F_score = \frac{2 * precision * recall}{precision + recall}$$

Where

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Ans 15)The fit method allows us to get the parameters of the scaling function. Ie Calculates the parameters(mean,standard deviation).

The transform() method will transform the dataset to proceed the for the further data analysis.Using the parameters we can transform the particular dataset.

The fit transform() will determine the parameters as well as transform the dataset.Actually it joins

the fit() and transform() for the transformation of the dataset.