Assignment No 7
MACHINE LEARNING

1. Which of the following in sk-learn library is used for hyper parameter tuning?
Ans : A) GridSearchCV() B) RandomizedCV()
A) GridSearchCV() B) RandomizedCV()
C) K-fold Cross Validation D) All of the above

2. In which of the below ensemble techniques trees are trained in parallel?
Ans : A) **Random forest**
A) **Random forest** B) Adaboost
C) Gradient Boosting D) All of the above

3. In machine learning, if in the below line of code:
*sklearn.svm.**SVC** (C=1.0, kernel='rbf', degree=3)*
we increasing the C hyper parameter, what will happen?
Ans: B) **The regularization will decrease**
A) The regularization will increase **B) The regularization will decrease**
C) No effect on regularization D) kernel will be changed to linear

4. Check the below line of code and answer the following questions:
*sklearn.tree.**DecisionTreeClassifier**(*criterion='gini',splitter='best',max_depth=None, min_samples_split=2)*
Which of the following is true regarding max_depth hyper parameter?
A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
B) It denotes the number of children a node can have.
C) both A & B
Ans : D) None of the above

5. Which of the following is true regarding Random Forests?
A) It's an ensemble of weak learners.
B) The component trees are trained in series
C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.
D)None of the above

6. What can be the disadvantage if the learning rate is very high in gradient descent?
A) Gradient Descent algorithm can diverge from the optimal solution.
B) Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle.
C) Both of them
D) None of them

7. As the model complexity increases, what will happen?
Ans : **B) Bias will decrease, Variance increase**
A) Bias will increase, Variance decrease
B) Bias will decrease, Variance increase
C) both bias and variance increase
D) Both bias and variance decrease.

8. Suppose I have a linear regression model which is performing as follows:
Train accuracy=0.95 and Test accuracy=0.75
Which of the following is true regarding the model?
A) model is underfitting B) model is overfitting
C) model is performing good D) None of the above

Q9 to Q15 are subjective answer type questions, Answer them briefly.
9. Suppose we have a dataset which have two classes A and B. The percentage of class A is 40% and percentage of class B is 60%. Calculate the Gini index and entropy of the dataset.
10. What are the advantages of Random Forests over Decision Tree?
Ans :

Decision Tree	**Random Forest**
A decision tree is a tree-like model of decisions along with possible outcomes in a diagram.	A classification algorithm consisting of many decision trees combined to get a more accurate result as compared to a single tree.
There is always a scope for overfitting, caused due to the presence of variance.	Random forest algorithm avoids and prevents overfitting by using multiple trees.
The results are not accurate.	This gives accurate and precise results.
Decision trees require low computation, thus reducing time to implement and carrying low accuracy.	This consumes more computation. The process of generation and analyzing is time-consuming.
It is easy to visualize. The only task is to fit the decision tree model.	This has complex visualization as it determines the pattern behind the data.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.
Ans : Remember when we used to solve maths problems in our childhood for mensuration where we had multiple units for the same quantity like oil which can take litres unit, ml units
So while solving problems we used to convert them into the same unit.
A similar concept applies to machine learning features also. It can be possible we can have multiple units and magnitudes which is varying. So the technique which helps in such conversion is known as feature scaling where we deal with their units and magnitudes.
Why is feature scaling required?
In features, we have 2 components magnitude i.e value and other is unit which gives information in which measure was measured. A dataset consists of many features with different magnitude and units.
Most of the machine learning algorithms use the distance between 2 data points for computation. If left alone, these algorithms only take in the magnitude of features neglecting the units. The results would vary greatly between different units, 5kg and 5000gms. Technically 5Kg and 5000Gms are the same. The features with high magnitudes will weigh in a lot more in the distance calculations than features with low magnitudes. This is hiding the original pattern of data. To suppress this effect, we need to bring all features to the same level of magnitudes. This can be achieved by scaling.
Which machine learning algorithms require scaling?

- **KNN and KMeans:-** It uses Euclidean distance hence scaling all numerical features to weigh equal.
- **PCA:-** PCA tries to get the features with maximum variance and the variance is high for high magnitude features. This skews the PCA towards high magnitude features.
- **Gradient Descent based Algorithms:-** Linear regression, Logistic regression, ANN tries to converge faster with scaled features.
- **Naïve Bayes and LDA:-** They internally handle the weightage of features so scaling may not have much effect.
- **Tree-Based Algorithms:-** DecisionTree, RandomForest, Boosting algorithms do not use distance for their computation. Hence scaling is not required.

What are different types of features scaling?

- **Standardization:-** It replaces values with their z scores

Xnew=(Xold-Xmean)/**σ**

This redistributes the features to their mean=0 and standard deviation=1. Its python implementation is available on the sklearn library.

- **Mean Normalization:-** This kind of scaling brings the distribution in the range of -1 to 1 with mean=0.

Xnew=(x-min(x)/max(x)-min(x)

Standardization and Mean Normalization is used for algorithms that assume zero centric data like PCA.

- **MinMax Scaler:-** This scaling technique brings the values in range 0 to 1.

Xnew=(x-min(x)/max(x)-min(x)

When to normalize and standardize the data?

Normalization is a good technique to use when you do not know the distribution of your data or when you know the distribution is not Gaussian (a bell curve). Normalization is useful when your data has varying scales and the algorithm you are using does not make assumptions about the distribution of your data, such as k-nearest neighbours and artificial neural networks.

Standardization assumes that your data has a Gaussian (bell curve) distribution. This does not strictly have to be true, but the technique is more effective if your attribute distribution is Gaussian. Standardization is useful when your data has varying scales and the algorithm you are using does make assumptions about your data having a Gaussian distribution, such as linear regression, logistic regression, and linear discriminant analysis.

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.

13. In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?

14. What is “f-score" metric? Write its mathematical formula.

Ans : The F-score, also called the F1-score, is a measure of a model’s accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into ‘positive’ or ‘negative’.

The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model’s precision and recall.

The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing.

It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.
The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

$$F_1 = \frac{2}{\frac{1}{\text{recall}} \times \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{\text{tp}}{\text{tp} + \frac{1}{2}(\text{fp} + \text{fn})}$$

Mathematical definition of the F-score

F-score Formula Symbols Explained

Precision	Precision is the fraction of true positive examples among the examples that the model classified as positive. In other words, the number of true positives divided by the number of false positives plus true positives.
Recall	Recall, also known as sensitivity, is the fraction of examples classified as positive, among the total number of positive examples. In other words, the number of true positives divided by the number of true positives plus false negatives.
tp	The number of true positives classified by the model.
fn	The number of false negatives classified by the model.
fp	The number of false positives classified by the model.

15. What is the difference between fit(), transform() and fit_transform()?

Ans : fit() Method

In the fit() method, we apply the necessary formula to the feature of the input data we want to change and compute the result before fitting the result to the transformer. We must use the .fit() method after the transformer object.
If the StandardScaler object sc is created, then applying the .fit() method will calculate the mean (μ) and the standard deviation (σ) of the particular feature F. We can use these parameters later for analysis.
Let's use the pre-processing transformer known as StandardScaler as an example and assume that we have to scale the features of self-created data. The example dataset in the code below is created using the arrange method and then divided into the training and testing datasets. After that, we create a StandardScaler instance and fit the feature of the training data to it to determine the mean and standard deviation to be utilized for scaling in the future.
The significance of separating the dataset into the train and test datasets before using any pre-processing process, such as scaling, must be emphasized.

Test data points represent real-world data. Therefore, we must only execute fit() to the training feature to prevent future data to our model.

transform() Method

To change the data, we most likely use the transform() function, where we perform the calculations from fit() to each value in feature F. We transform the fit computations. Hence we must use .transform() after we have applied the fit object.

When we make an object using the fit method, we utilize the example from the section above and place the object in front of the.

The scale of the data points is transformed using the transform and fit_transform method, and the output we receive is always a sparse matrix or array.

fit_transform() Method

The training data is scaled, and its scaling parameters are determined by applying a fit_transform() to the training data. The model we created, in this case, will discover the mean and variance of the characteristics in the training set.

The mean and variance of every feature reported in our data are calculated using the fit approach. The transform method transforms all features using the corresponding means and variances.

We wish scaling to be implemented in our testing data, but we also don't want our model to be biased. We expect our test set of data to be entirely fresh and unexpected for our model. In this situation, the convert approach is useful.

WORKSHEET 7 SQL

Q1 and Q2 have one or more correct answer. Choose all the correct option to answer your question.

1. The primary key is selected from the
A. Composite keys
B. Candidate keys
C. Foreign keys
D. Determinants

2. Which is/are correct statements about primary key of a table?
A. Primary keys can contain NULL values.
B. Primary keys cannot contain NULL values...
C. A table can have only one primary key with single or multiple fields....
D. A table can have multiple primary keys with single or multiple fields.

Q3 to Q10 have only one correct answer. Choose the correct option to answer your question.

3. Which SQL command is used to insert a row in a table?
A. Select
B. Create
C. Insert
D. Drop

4. Which one of the following sorts rows in SQL?
A. SORTBY
B. ALIGNBY
C. ORDERBY
D. GROUPBY

5. The SQL statement that queries or reads data from a table is
A. QUERY
B. READ
C. SELECT
D. QUERY

6. Which normal form is considered adequate for relational database design?
A. 1NF
B. 2NF
C. 3NF
D. 4NF

7. SQL can be used to
A. Create database structures only
B. Modify database data only
C. All of the above can be done by SQL
D. Query database data only

Q11 to Q15 are subjective answer type questions, Answer them briefly.

11. What are joins in SQL?
Ans : A SQL join tells the database to combine columns from different tables. We normally join tables by matching the foreign keys in one table to the primary keys in another.

For example, every record in the **products** table has a unique ID in the **products.id** field: that's the primary key. To match the key, every record in **orders** has a product ID in the **orders.product_id** field: that's a foreign key. If we want to combine information about an order with information about the product that was ordered

12. What are the different types of joins in SQL?

Ans: Here are the different types of the JOINs in SQL:

- (INNER) JOIN: Returns records that have matching values in both tables
- LEFT (OUTER) JOIN: Returns all records from the left table, and the matched records from the right table
- RIGHT (OUTER) JOIN: Returns all records from the right table, and the matched records from the left table
- FULL (OUTER) JOIN: Returns all records when there is a match in either left or right table

13. What is SQL Server?

Ans : SQL Server is a relational database management system, or RDBMS, developed and marketed by Microsoft.

Similar to other RDBMS software, SQL Server is built on top of SQL, a standard programming language for interacting with relational databases. SQL Server is tied to Transact-SQL, or T-SQL, the Microsoft's implementation of SQL that adds a set of proprietary programming constructs.

SQL Server works exclusively on the Windows environment for more than 20 years. In 2016, Microsoft made it available on Linux. SQL Server 2017 became generally available in October 2016 that ran on both Windows and Linux.

14. What is primary key in SQL?

Ans : The PRIMARY KEY constraint uniquely identifies each record in a table. Primary keys must contain UNIQUE values, and cannot contain NULL values. A table can have only ONE primary key; and in the table, this primary key can consist of single or multiple columns (fields).

A primary key is a field in a table which uniquely identifies each row/record in a database table. Primary keys must contain unique values. A primary key column cannot have NULL values.

A table can have only one primary key, which may consist of single or multiple fields. When multiple fields are used as a primary key, they are called a composite key.

If a table has a primary key defined on any field(s), then you cannot have two records having the same value of that field(s).

15. What is ETL in SQL?

Ans : ETL, which stands for "extract, transform, load," are the three processes that, in combination, move data from one database, multiple databases, or other sources to a unified repository—typically a data warehouse. It enables data analysis to provide actionable business information, effectively preparing data for analysis and business intelligence processes.

As data engineers are experts at making data ready for consumption by working with multiple systems and tools, data engineering encompasses ETL. Data engineering involves ingesting, transforming, delivering, and sharing data for analysis. These fundamental tasks are completed via data pipelines that automate the process in a repeatable way. A data pipeline is a set of data-processing elements that move data from source to destination, and often from one format (raw) to another (analytics-ready).

STATISTICS WORKSHEET-7

1. A die is thrown 1402 times. The frequencies for the outcomes 1, 2, 3, 4, 5 and 6 are given in the following table:

Outcome	1	2	3	4	5	6
Frequency	400	300	157	180	175	190

Find the probability of getting 6 as outcome:

a) 0.34

Ans : b) 0.135

c) 0.45

d) 0.78

Solution Probability of an Event E can be given as,

P(E)=Number of occurrences of Event E/Total number of outcomes

Here, total number of outcomes =1402

We can find the number of outcomes in the given frequency table.

So, Probability of outcome 6, P(E6)=190/1402 = 0.135

Probability of outcome 5, P(E5)=175/1402 = 0.124

Probability of outcome 4, P(E4)=180/1402 = 0.128

2. A telephone directory page has 400 telephone numbers. The frequency distribution of their unit place digit (for example, in the number 25827689, the unit place digit is 9 is given in table below:

First row refers to the digits .

Second row to their frequencies.

0	1	2	3	4	5	6	7	8	9
44	52	44	44	40	20	28	56	32	40

What will be the probability of getting a digit with unit place digit odd number that is 1, 3,5,7,9?

a) 0.67

b) 0.60

c) 0.45

d) 0.53

Ans =52+44+20+56+40/400=0.53

3. A tyre manufacturing company which keeps a record of the distance covered before a tyre needed to be replaced. The table below shows the results of 1100 cases.

Distance (miles)	<4000	4000-9000	9001-14000	>14000
Frequency	20	260	375	445

If we buy a new tyre of this company, what is the probability that the tyre will last more than 9000 miles?

a) 0.67

b) 0.459

c) 0.745

d) 0.73

solution : **375+445/1100=0.745**

4. Please refer to the case and table given in the question No. 3 and determine what is the probability that if we buy a new tyre then it will last in the interval [4000-14000] miles?
a) 0.56
b) 0.577
c) 0.745
d) 0.73
solution : 260+375/1100=0.5772

5. We have a box containing cards numbered from 0 to 9. We draw a card randomly from the box. If it is told to you that the card drawn is greater than 4 what is the probability that the card is odd?
a) 0.5
b) 0.8
c) 0.6
d) 0.7
solution : **Step-by-step explanation:**
⇒ Total possible outcomes = 10
⇒ No. of Favorable outcomes = 1, 3, 5 , 7 , 9 ⇒ 5
Therefore,
Probability = 5/10 = **0.5**
Probability P(F) =(Total no.of possible outcomes)/(No.of favorable outcomes)

6. We have a box containing cards numbered from 1 to 8. We draw a card randomly from the box. If it is told to you that the card drawn is less than 4 what is the probability that the card is even?
a) 0.33
b) 0.40
c) 0.56
d) 0.89

7. A die is thrown twice and the sum of the numbers appearing is observed to be 7. What is the conditional probability that the number 6 has appeared at least on one of the die?
a) 0.45
b) 0.37
c) 0.33
d) 0.89

Solution:
A= sum of the numbers appearing on two dice is 7
(1,6), (2,5),(3,4),(4,3),(5,2),(6,1)
n(A) =6
B= number 6 has appeared at least once
=(1,6),(6,1),(2,6),(6,2),(3,6),(6,3),(6,6),(6,5),(5,6),(6,6),(6,4),(4,6)
A∩B= (1,6), (6,1)
n(A∩B)=2
Required probability = P(B/A)= n(A∩B)/ n(A)=2/6

8. Consider the experiment of tossing a coin. If the coin shows tail, toss it again but if it shows head, then throw a die. Find the conditional probability of the event that 'the die shows a number greater than 4' given that 'there is at least one Head'.

a) 0.1

b) 0.22

c) 0.38

d) 0.45

The sample space of an experiment that there is at least one tail and the die is thrown can be given as,

$S = (H , H) , (H , T) , (T , 1) , (T , 2)(T , 3)(T , 4)(T , 5)(T , 6)$

Where (T , i) denotes that the first toss resulted in a tail and the number i appeared on the die for $i = 1 , 2 , 3 , 4 , 5 , 6$.

Now, the probabilities of the above mentioned 8 elementary events are $1/4 , 1/4 , 1/12 , 1/12 , 1/12 , 1/12 , 1/12 , 1/12$ respectively.

Now, let the event that 'there is at least one tail' be represented by A and the event 'the die shows a number greater than 4' be represented by B . Then,

$A = (H , T) , (T , 1) , (T , 2) , (T , 3) , (T , 4) , (T , 5) , (T , 6)$

And, $B = (T , 5) , (T , 6)$

So, $A \cap B = (T , 5) , (T , 6)$

Now, $P(A) = P(H , T) + P(T , 1) + P(T , 2) + P(T , 3) + P(T , 4) + P(T , 5) + P(T , 6)$

$\Rightarrow P(A) = 1/4+1/12+1/12+1/12+1/12+1/12+1/12$

$\Rightarrow P(A) = 3+1+1+1+1+1+1/12$

$\Rightarrow P(A) = 9/12$

$\Rightarrow P(A) = 3/4$

And, $P(A \cap B) = P(T , 5) + P(T , 6)$

$\Rightarrow P(A \cap B) = 1/12+1/12$

$\Rightarrow P(A \cap B) = 1+112$

$\Rightarrow P(A \cap B) = 2/12$

$\Rightarrow P(A \cap B) = 1/6$

Now, as it is already known,

$P(B| A) = P(A \cap B)P(A)$

$\therefore P(B| A) = 1/6*3/4$

$\Rightarrow P(B|A) = 1/6 \cdot 4/3$

$\Rightarrow P(B|A) = 2/9$

Hence, the conditional probability of the event that the die shows a number greater than 4 given that there is at least one tail is $2/9$ or 0.22.

9. There are three persons Evan, Ross and Michelle. These people lined up randomly for a picture. What is the probability of Ross being at one of the ends of the line?
a) 0.66
b) 0.45
c) 0.23
d) 0.56

10. Let us make an assumption that each born child is equally likely to be a boy or a girl. Now suppose, if a family has two children, what is the conditional probability that both are girls given that at least one of them is a girl?
a) 0.33
b) 0.45
c) 0.56
d) 0.26

11. Consider the same case as in the question no. 10. It is given that elder child is a boy. What is the conditional probability that both children are boys?
a) 0.33
b) 0.23
c) 0.5
d) 0.76

12. We toss a coin. If we get head, we toss a coin again and if we get tail we throw a die. What is the probability of getting a number greater than 4 on die?
a) 0.166
b) 0.34
c) 0.78
d) 0.34

13. We toss a coin. If we get head, we toss a coin again and if we get tail we throw a die. What is the probability of getting an odd number on die?
a) 0.345
b) 0.79
c) 0.2
d) 0.25

14. Suppose we throw two dice together. What is the conditional probability of getting sum of two numbers found on the two die after throwing is less than 4, provided that the two numbers found on the two die are different?
a) 0.3
b) 0.56
c) 0.24
d) 0.06

15. A box contains three coins: two regular coins and one fake two-headed coin, you pick a coin at random and toss it. What is the probability that it lands heads up?

a) 1/3
b) 2/3
c) 1/2
d) 3/4