1. Deep learning (NN, RNN, LSTM, CNN etc…) with real time in python packages(numpy, tensorflow, keras, scikit learn, torch)
2. Machine learning algorithms
3. Statistics
4. Time series Forecasting
5. Time series Forecasting in Deep learning
6. Python API creation
7. NLP
8. Mongo DB
9. Timeseries DB
10. AWS
11. AZURE
12. GCP
13. Docker image creation
14. Kubernetes
15. GO

Machine Learning,

1. Decision Tree
2. Random Forest
3. SVM
4. Bayesian interface
5. Adboost
6. Xgboost
7. Markov chain
8. Monte Carlo simulations
9. ADDA
10. Ada Match
11. Self ensemble
12. Cross Validation

Statistics,

1. Pdf

Answer: probability for a continuous random variable is given by areas under pdf's

1. Cdf

Answer: Integration of pdf is cdf

1. Why p-value should be less than 5%?

Answer: If p-value is less than 5%, it means the data used for the model is greater than 95% of data. Let’s consider if we fit the features to the algorithm if we get one feature as p-value 0.38. This doesn’t contribute to model. Reason being, for that feature only 62% of the data getting used in model. Rest 38% of the data it doesn’t use. So, we remove that feature.

1. Contingency table

Answer: It is the tabular representation of the categorical data. Contingency table shows the frequency distribution of combinations of values of two discrete variable x and y.

Table

Description automatically generated

1. Degrees of freedom

Answer: In the most of formulae we are using n-1 in denominator. The reason for that could be if we use n then there will be slightly biased. But we do n-1 in denominator the bias will be reduced. However, it doesn’t matter much if the dataset is huge. And the n-1 is called as degrees of freedom.

1. Maximum likelihood estimation

Answer: If we have data with mean and standard deviation (sigma), the future data which is similar with current data will have same mean and standard deviation (sigma) of current data. So that it will increase the likelihood of estimation.

1. Central limit theorem

Answer: The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.

1. Bootstrap

Answer: A sample taken with replacement from an observed dataset. The sample may not be in normal distribution.

1. Permutation test

Answer: The process of combining two groups of samples and randomly creating the sample from the combined group.

1. Resampling

Answer: Process of taking repeated samples from observed data. Repeated samples taking procedure can be bootstrap or permutation test.

1. Standardization or Normalization

Answer: Subtract each data point with its mean and divide it by standard deviation will make the data as Standardization or Normalization

1. Z distribution

Answer: It is nothing but the Standardization.

1. T distribution

Answer: It is like Z test, but here it is sample standard distribution. However, Z test is population standard distribution. T test will have longer tails. Shape of the distribution looks like T shape. T distribution with high degrees of freedom will be exactly same as Z distribution.

1. One tail test
2. Two tail test
3. Binomial distribution

Answer: If the outcome of the data is yes/no or True/False. Then it is in binomial distribution.

1. Poison distribution
2. Exponential distribution
3. Chisquare test- It’s the procedure to test the observed count is as expected count which shows the independence. Like for example buy a particular item is independent of gender.
4. Fisher test- It is also same as Chisquare test to check independence. But it works only smaller dataset.
5. A/B testing – An A/B test is an experiment with two groups to establish which of two treatments, products, procedures, or the like is superior.
6. Stationary poison process
7. Stochastic proximity embedding (SPE)

**Key Points**

1. In the most of formulae we are using n-1 in denominator. The reason for that could be if we use n then there will be slightly biased. But we do n-1 in denominator the bias will be reduced. However, it doesn’t matter much if the dataset is huge. And the n-1 is called as degrees of freedom

Mathamatics,

1. Where does Fourier transform fit in ML models?
2. Wallet packet decomposition