

Plan of Action

Introduction:

- Introduction to python basic
- List
- Dictionary
- Tuples
- Map, reduce, lambda, filter
- Recursion
- Graphs
- Algos
- Pandas
- Numpy

Webinar 1: introduction and applying numpy and pandas to titanic data set.

visualisation:

- .2d plot, pie plot, scatter plot, bar charts, basic plotting
- Cdf, pdf – distribution plot
- Percentiles
- Box plot, violin plots

Webinar 2 visualising iris data set.

Statistics:

- Gaussian distribution and other distributions
- Kernel density estimation
- Central limit theorem
- Q-q plot
- Chebyshev inequality
- Log normal distribution
- Power law
- Box cox and log log plot
- Covariance and correlation
- Pearson correlation coefficient
- Spearman rank correlation
- Confidence interval
- Empirical bootstrap
- Student t distribution
- Null hypothesis
- Permutation testing
- Ks test
- Chi square test
- Anova test
- Proportional sampling

Webinar3: discussion on
fitting to gaussian values

Dimensionality reduction:

- pca.
- Tsne
- lca

Webinar 4 feature engineering and gaussian fitting techniques all on titanic

knn:

- Types of distance in statistics
- Intuition about knn
- Choosing appropriate value of k
- K fold cross validation
- Overfitting and under fitting
- Multiclass classification
- Knn for regression
- Voroni diagram
- Building a kd tree
- Hashing and lsh
- Training and testing distribution
- Impact of outliers
- Local outlier factor

Webinar 5 knn on titanic and iris

Feature engineering revisited:

- Impact of outliers
- Curse of dimensionality
- Imbalanced dataset
- Parameters for scoring

Webinar 6 all techniques to handle outliers

Webinar 7 all techniques to handle missing values

Webinar 8 all techniques to handle imbalanced data set

Webinar 9 all techniques for feature selection

Webinar 10 all techniques for hyper parameter optimization

Webinar 11 all techniques for categorical values

Webinar 12 all techniques for cross validation

Linear regression:

- Intuition
- Linear regression assumptions
- R squared and adjusted r square
- Ridge and lasso regression

Webinar 13 advanced house price prediction

Logistic regression:

- Intuition
- Sigmoid function squashing
- Log loss

Webinar medical data set logistic regression

Decision Tree:

- Intuition
- Entropy and information gain
- Gini
- Cart, chaid, c4.5
- Numerical data set
- Problem with decision tree

Webinar project

Ensembling Techniques:

- Intuition
- Bootstrap method
- Random forest
- Random forest regressor

Webinar project

Boosting techniques:

- ADA boost
- XGboost
- variations

Webinar project

Clustering techniques:

- Intuition
- Kmeans, k mediodes, k medians
- dbscan
- hierarchical

Iske aage lagta bhi nahi mujhe ho bhi payegga agar time hua toh sirf projects karenge
Next batch can continue after this