Machine learning:

1. Regression: Output label is numerical ====

* Linear regression ( OLS) differentiation

======= Regularization ============

* Lasso Regression ( need to cover)
* Ridge regression (need to cover)

1. Classification: Output label is categorical

* Logistic regression ( Sigmoid function)
* Decision Tree (Entropy and Gini)
* Naïve bayes (Probability multiplication)
* KNN (Distance metric)

=====================================

* SVM
* Bagging Boosting
* Recommendation engine

1. Cluster : Output label is not present

* K-means (Distance )
* Hierarchical (Linkage functions)

Important topics:

* PCA : Principle component analysis
* Bias -variance
* Overfit- underfit-normal fit
* Optimization : How to avoid overfitting
* Regularization:
* Feature selection methods
* Hyper parameter tuning

Metrics:

* Regression metrics
* Classification metrics
* Cluster metrics

DS : ML DL NLP

DL : 1st week

NLP : analysis

K-means

* K= number of groups hyper parameter
* Will calculate distance between observations with group of means
* Which ever is the lowest distance that observation will go into that group
* When observations are updated in a group, mean will change
* Will repeat this process till the means are fixed

KNN

* K= number of neighbours hyper parameter
* Calculate the distance between observations
* In that one observation is test sample
* Another observations are neighbours
* Majority vote

Both are using Distance metrics

Hierarchical:

1. First it will consider every observation is one group

|  | A | B | C | D | E |
| --- | --- | --- | --- | --- | --- |
| A | - | 2 | 7 | 9 | 4 |
| B | 2 | - | 9 | 11 | 14 |
| C | 7 | 9 | - | 4 | 8 |
| D | 9 | 11 | 4 | - | 2 |
| E | 4 | 14 | 8 | 2 |  |

|  | A | C | D | EB |
| --- | --- | --- | --- | --- |
| A | - | 7 | 9 |  |
| C | 7 | - | 4 |  |
| D | 9 | 4 | - |  |
| EB |  |  |  | - |

Simple linkage:

|  | A | C | D | EB |
| --- | --- | --- | --- | --- |
| A | - | 7 | 9 | 4 |
| C | 7 | - | 4 | 9 |
| D | 9 | 4 | - | 11 |
| EB | 4 | 9 | 11 | - |

|  | A | C | BED |
| --- | --- | --- | --- |
| A | - | 7 |  |
| C | 7 | - |  |
| BED |  |  | - |

|  | A | C | BED |
| --- | --- | --- | --- |
| A | - | 7 | 9 |
| C | 7 | - | 9 |
| BED | 9 | 9 | - |

