PHASE-1

Machine Learning Techniques

For each topic below, use the resources (links) to learn about them. With each task, a dataset from the UCI ML Repository will be given, do some exploratory analysis using pandas, seaborn, matplotlib, numpy and use the methods learnt on the data using functions from sklearn library and learn hands-on about the ML methods you learnt.

Common Resources:

- ML in a Nutshell
- Krish Naik

Linear Regression: (3 days)
- Week 1 of the Machine Learning course on Coursera
- Week 2 of the Machine Learning course on Coursera

Mini Assignment: Exploratory data analysis and applying Linear regression on this dataset

Logistic Regression: (3 days)
- Week 3 of the Machine Learning course on Coursera

Introduction to Neural Networks: (4 days)
- Deep Learning specialization| Course1 | Neural Network and Deep Learning |
  - Part 1 of Week 2 will be kind of repetitive, teaching Logistic regression. So feel free to skip those videos to save time
  - (You can skip Interviews or you can watch them after completing the course)
  - Attempt Assignments nicely, if you think your code is specific to the test cases, we can check them for you :)
- Neural Networks Series by 3Blue1Brown (Do this in parallel with week 1 of above course)
  - If you feel that you understand the things described in the above material thoroughly, you can skip this and move on to more advanced material later on.
  - Consists of 4 episodes
  - Gives an extremely good intuitive (although basic) understanding of Neural Networks
  - A total of ~64 minutes of (brilliant) content
  - Worth watching more than once to clear up doubts later on
  - Your objective is to understand intuition. Don't try to memorize the math. Asas long as you're able to understand everything Grant is saying, that is enough.
  - However, make sure that you understand everything in these videos.
  - Mantra: Pause and Ponder

MINI PROJECT 1: (3 days)
- Implement a linear regressor for the given dataset from scratch only numpy
- You CANNOT use sklearn for the first 5 tasks.
- See the attached pdf and the code.py file where u need to implement the functions and get output (hints are given in code.py) for the PS and do the coding accordingly.
- The PS is an actual assignment from a course in insti (but it is quite easy so don't worry), so ignore instructions about the Kaggle contest and report stuff. Focus solely on the code.py file.
- What you learn out of this: numpy and pandas skills, googling and documentation reading skills.
- All required files to complete are in the link below.
- PS Link Here

K-Means: (3 days)
- Week 8 of the Machine Learning course on Coursera
- Guide to KMEANS Article

Mini assignment: Try KMeans on this dataset - customer segmentation dataset. Divide the customers into various clusters

  For clearer understanding:
  - K-Means Krish Naik
  - StatQuest

KNN: (2 days)
- https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/
- StatQuest: K-nearest neighbors, Clearly Explained
- K Nearest Neighbor classification with Intuition and practical solution (optional)

Mini assignment:
A. Try KNN on this dataset - customer segmentation dataset and give your conclusions.
B. How do you find the optimal value of k? Research on methods employed and implement it and infer best k on the iris dataset.

PCA: (4 days)
- https://www.youtube.com/watch?v=HMOI_lkzW08&ab_channel=StatQuestwithJoshStarmer
- https://www.youtube.com/watch?v=FgakZw6K1QQ&ab_channel=StatQuestwithJoshStarmer
- Go through the Dimensionality Reduction section of Week 8 of the Machine Learning course on Coursera
- https://youtu.be/QdBy02ExhGl

MINI PROJECT 2: (2 days)
- Load MNIST dataset into the memory in colab.
- Pick your lucky digit and extract the images of that particular digit. Here on, we will restrict to that particular digit.
- those 28 x 28 images and perform dimensionality reduction using PCA on it. (Do not forget to subtract the mean). Give a reason for the choice of number of dimensions to project on.

- Take any image from the dataset, get its representation in the hyperplane. Reconstruct the high dimensional image from the low dimensional representation and show it alongside the original image.
- Suggest a method to generate a new image of that particular digit
(Hint : Use the svd decomposition that you have in your hand)
- (Fun stuff). Take a look into the 3 images mu, mu-root(lambda)v, mu+root(lambda)v where mu is the mean, lambda is the largest eigenvalue and v is the corresponding eigenvector. Can you guess what they would look like ? ;)

(You are allowed to use scipy/ sklearn's functions to complete this task from obviously numpy, pandas and matplotlib.)

Decision Trees: (just basic intuition needed)(1 day)
- https://www.geeksforgeeks.org/decision-tree/
- https://www.youtube.com/watch?v=7VeUPuFGJHk&ab_channel=StatQuestwithJoshStarmer
- https://www.youtube.com/watch?v=ot75kOmpYjI&ab_channel=KrishNaik

Random Forest Regressors & Classifiers: (just basic intuition needed)(1 day)
- https://www.youtube.com/watch?v=v6VJ2RO66Ag
- https://youtu.be/J4Wdy0Wc_xQ
- https://towardsdatascience.com/understanding-random-forest-58381e0602d2

If you need additional assignments, given that you have time we can send you more resources