GUIDE TO PREPARE FOR LAB EXAM ASAP

- Revise python skills(numpy ,pandas, dictionary), should clearly understand how to process data if a csv file is given. Should be able to process data into X(input) and y(output set), remove features, convert to jumpy array etc.

- Learn how we can generate sample datasets for regression, classification etc using inbuilt functions if dataset is not given. We had seen sample examples towards the end. See how it is done for both regression and classification task.

- Learn to identify important features using heat map.
- Split data into training and test set.
- Learn basic plotting skills using matplotlib

- Implement Gradient Descent

- For a given dataset or dataset generated perform (possible Q WOULD BE TO COMPARE ACCURACY OF DIFFERENT IMPLEMENTATIONS)-

1. linear regression,
2. multivariate regression and
3. polynomial regression.
4. See how results can be improved using Regularisation - LASSO and RIDGE
5. print accuracy, SSE, RMSE scores for each model and compare results.
6. Perform Regression using decision trees and CART
7. Perform regression using Gradient Boosting

- For a given dataset or dataset generated perform (possible Q WOULD BE TO COMPARE ACCURACY OF DIFFERENT IMPLEMENTATIONS in a single program) -

8. logistic regression (binary, softmax),
9. linear SVM,
10. different Kernel SVMs - radial basis, gaussian, linear, polynomial etc
11. SVM - soft margin by fine tuning parameters like C
12. Linear classification using perceptron model
13. Naive Bayes using different density functions - gaussian Bernoulli, multinomial etc
14. K NN , adjusting values of k, Parzen window
15. Decision tree - using gini index as split criteria. how we can specify/adjust depth of tree
16. Printing evaluation parameters - confusion matrix, Score summary(presion, recall, F1, accuracy) and plotting ROC curve
17. Bagging using decision tree vs single decision tree accuracy comparison
18. Bagging using different Base estimators like SVC etc
19. Random Forest - see how we specify depth, no of estimators
20. Comparing different classifiers individually and combining them together using Ensemble VOTING and see accuracy
21. How we can combine different models using VOTING
22. ADABOOST using diff base estimators
23. Checking accuracy using cross validation.

- Reduce dimensionality using PCA- see how we find Eigen values, vectors, decide dimension , print transformed matrix

- Generate dataset/csv file - perform clustering - K means (decide k by iteration), Hierarchical clustering (single, maximum, average linkage), density based DIANA. -

- Go through all maths codes especially generating different distributions, gradient descent, linear programming, minimising a function with respect to another.