

CLASS - Portions

- Week1-Introduction to Python, NumPy Pandas
- Week2-Introduction to SciPy, Matplotlib and Seaborn, Py-torch
- Week3-Linear Regression Dataset splitting
 - Multiple Regression - OLS (sklearn)
 - Metrics for comparison
 - R2
 - MAE
 - MSE
 - RMSE
 - LR with SKlearn
 - Cross Validate
 - K - FOLD
 - Leave One Out
- Week4-Linear Regression Diagnostics PCR PLS Notes
 - LR Diagnostics - sklearn and stats model (verifying model assumptions)
 - Residual Normality check
 - QQ plot
 - Histogram
 - Homoscedastic - Standardised residual [plot + hypothesis]
 - Correlation of residuals with features (Pearson-r)
 - Auto-Correlation (ACF)
 - PCR and PLS
 - PCR = PCA + LR with pipeline
 - PLS
- Week5-Bias Variance Tradeoff,Regularization
 - Data structure - Nonlinear - add polynomial features - overfit the data
 - Define own pipeline - poly + LR
 - MSE with degrees - Train and test
 - Ridge regression
 - reduce overfitting with overfitted degree - L2 regularisation (Coeff reduces)
 - Plot Coeff
 - Lasso - Laplacian
 - Plot Coeff - important feature selection
- Week6-Dimensionality Reduction Techniques
 - PCA
 - NMF - Non Negative Matrix factorisation
 - t-SNE - t-distributed Stochastic Neighbour Embedding
 - UMAP - Uniform Manifold Approximation and Projection
- Week7- Classification methods- Logistic, NB, LDA, QDA, KNN
 - Correlation Heat Map - feature selection
 - Label Encoder - Encoding class names in numerical value
 - Confusion Matrix/Classification report
 - Logistic regression
 - Gaussian Naïve Bayes
 - LDA
 - QDA
 - KNN
 - ROC Curve - binary and multiclass (One vs Rest class)
- Week8-Decision Tree and Its Variants
 - Making different datasets - moons, blob etc
 - Decision Tree - Interpretable and High Variance Model

- Logistic Regression vs Decision Tree Classifier
- Hyper Parameter tuning - Grid Search CV
- Bootstrap aggression or Bagging Classifier
- Random Forest Classifier
- Gradient Boosting Classifier
- Comparing Models Performance
- Comparison of each tree based algorithm for mobile price prediction
- Plot - Feature importance
- Week9-SVM
 - SVM
 - Hyper Parameter Tuning - c, Kernel and gamma
 - One Hot Encoding
 - Confusion Matrix and Classification report
- Week10-Clustering
 - Blob, moon, Circle dataset
 - K-Means
 - Score - Rand Score, Silhouette Score
 - Hierarchical Clustering
 - Linkage , Dendogram
 - Spectral Clustering
- Week11-SVR and Deep Learning
 - SVR with Linear Kernel, rbf etc
 - NN - ANN/MLP - MNist classification
- Week12_IndustrialAI CNN RNN
 - Image Processing using Kernel - Blur, Edge Detection, Sharpen
 - Image Classification
 - Visualising the output at each layer

ASSIGNMENTS

- Assignment 1
 - Benign, Malignant Cancer
 - Correlation based feature selection
 - Correlation Map - Multicollinearity
 - Outbreaks in India
 - Word Cloud
 - Geospatial plot (geopandas)
 - Scatter and pie
- Assignment 2
 - LR Diagnostics
 - Multivariate LR model
 - PCR, PLS - Compare performance
 - Optimal number of PCs selection
- Assignment 3
 - Ridge
 - Lasso
 - Effect of regularisation term alpha on Coeff's and performance
 - Compare
- Assignment 4
 - Logistic, LDA, QDA
 - Decision Tree, Hyper Parameter tuning
 - Bagging, Random Forest
 - Compare
- Assignment 5
 - K Means

- Hierarchical Clustering - effect of linkage and distance metric on number of clusters obtained
 - Spectral - verify if its better
 - UMAP and t-SNE
- Assignment 6
 - SVM - ROC-AUC Metric
 - NN - ANN
 - Reduce Overfitting - drop out/early call back
- Assignment 7
 - CNN - balance dataset (weighted sampling or weighted loss or augment)
 - Prediction accuracy, precision, recall and F1 score
 - Use Pretrained model and compare