CosmiCode

Educating the Next Wave of Tech Pioneers.

Week 1 Tasks

- **Task 1:** Implement a basic linear regression model from scratch using NumPy. Fit the model to a simple dataset and plot the regression line.

- **Task 2:** Perform data cleaning on a real-world dataset. Handle missing values, detect and remove outliers, and normalize/standardize the data using pandas.

- **Task 3:** Implement feature selection using correlation matrix and mutual information. Visualize the important features.

- **Task 4:** Conduct exploratory data analysis (EDA) on a dataset. Generate insightful visualizations using matplotlib and seaborn.

- **Task 5:** Apply PCA (Principal Component Analysis) to reduce the dimensionality of a dataset and visualize the results in 2D/3D.

Multan, Pakistan

+92 304 5892020

cosmicodepk@gmail.com

CosmiCode

Educating the Next Wave of Tech Pioneers.

Week 1 Guide

Objective: Understand the basics of machine learning, data preprocessing techniques, and tools.

1. Task 1: Implement a basic linear regression model from scratch using NumPy. Fit the model to a simple dataset and plot the regression line.
   - Guide: Learn the basics of linear regression and NumPy. Use gradient descent for parameter estimation. Visualize the results using matplotlib.
2. Task 2: Perform data cleaning on a real-world dataset. Handle missing values, detect and remove outliers, and normalize/standardize the data using pandas.
   - Guide: Use pandas for data manipulation. Handle missing values with imputation methods, detect outliers using statistical techniques, and apply normalization/standardization.
3. Task 3: Implement feature selection using correlation matrix and mutual information. Visualize the important features.
   - Guide: Calculate the correlation matrix to find correlated features. Use mutual information for feature selection. Visualize feature importance using bar plots.
4. Task 4: Conduct exploratory data analysis (EDA) on a dataset. Generate insightful visualizations using matplotlib and seaborn.
   - Guide: Perform univariate and multivariate analysis. Create histograms, box plots, scatter plots, and pair plots to understand the data distribution and relationships.
5. Task 5: Apply PCA (Principal Component Analysis) to reduce the dimensionality of a dataset and visualize the results in 2D/3D.
   - Guide: Understand the principles of PCA. Use scikit-learn to apply PCA on the dataset. Visualize the principal components using 2D/3D scatter plots.

Multan, Pakistan

+92 304 5892020

cosmicodepk@gmail.com