Linear Algebra and Probability

In this assignment, you will apply Linear Algebra and Probability techniques to analyze the Wine Quality Dataset. The dataset consists of 1599 rows and 12 columns, detailing various chemical properties of red and white wines and their quality scores. Your tasks will focus on vectors, matrices, eigenvalues, eigenvectors, and basic probability concepts. You can download the dataset from the link given below. The dataset contains 1599 samples of wine, with features such as acidity, alcohol content, pH, residual sugar, and others, alongside the target variable, wine quality.

1. Load the dataset and handle any missing data by replacing null values with the mean value of the respective column. (Score 2)
2. Extract the following columns as vectors: alcohol,citric acid. (Score 2)
3. Select two features (e.g., alcohol and density) from the dataset and calculate the covariance matrix using np.cov(X.T), where X is the feature matrix consisting of the selected columns.(Score 2)
4. Perform eigen decomposition on the covariance matrix you computed in question 3. Identify and interpret the results:Identify the top 2 eigenvalues of the covariance matrix,Identify the corresponding eigenvectors.(Score 2)
5. Which wine quality is most common in the dataset? How can you interpret the distribution of wine quality scores? (Score 2)

https://drive.google.com/file/d/1OnJ8qjX8HAlNU5JCHFRdYjMHLd78Ft56/view?usp=drive_link