**Assignment (20 Marks)**

Q.1 Marks 4

In the context of machine learning, eigenvalues and eigenvectors are crucial concepts, particularly in methods like Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and various linear transformations in feature space. These mathematical tools help in understanding and reducing dimensionality, improving algorithm performance, and extracting meaningful patterns from the data.

Consider a scenario where a data scientist is working on a large dataset to predict customer churn. To simplify the predictive modeling, the scientist wants to reduce the dataset's dimensionality without losing significant information. This is often achieved through PCA, which requires computation of eigenvalues and eigenvectors.

To grasp the underlying concepts, the scientist starts with a simpler problem — computing the eigenvalues of a smaller matrix. This matrix could represent a simplified transformation in a reduced feature space. Here is the matrix in question:

1. **Calculate Eigenvalues**: The objective is to compute eigenvalues of matrix A. This will provide insights into the variance explained by the components when matrix A is used in PCA.
2. **Interpret Results**: Understanding the eigenvalues helps in determining the importance of each principal component. What is the relation between the sum of the Eigenvalues with matrix A? What is the relation between product of the Eigenvalues with matrix A?

Q.2 Marks 4

As a senior data analyst at XYZ Corp, a leader in retail, you've been tasked with evaluating the efficacy of a recently concluded, multifaceted campaign. This campaign, aimed at boosting the sales of a high-profile product, spanned multiple channels including digital (emails and social media), traditional media (TV and print), and various experimental platforms.

The primary objective is to assess the performance of a predictive model designed to identify potential buyers among those exposed to the campaign. This assessment is crucial, as its outcomes will guide the strategic planning for upcoming marketing efforts.

The dataset collected is extensive, including:

* Engagement metrics such as number of ad impressions, clicks, and time spent on product pages.
* Viewer demographics including age, location, and previous purchase history.
* Follow-up actions, specifically whether viewers purchased the product post-ad exposure.
* Additional engagement metrics unrelated to the campaign, like overall website traffic and user interaction across non-campaign content.

Extended Data Description:

* From the campaign, 1,000 viewers who clicked on the ads were randomly selected for detailed tracking.
* The predictive model used in the campaign anticipated that 350 of these tracked viewers would purchase the product.
* Post-campaign analysis confirmed that 200 of these viewers actually purchased the product, suggesting a need to evaluate the predictive accuracy of the model.
* The model's specific predictions indicated that out of the 350 viewers it identified as likely purchasers, only 200 completed a purchase. This indicates potential overfitting or misalignment with actual consumer behavior.
* Furthermore, the model identified 500 viewers who would not purchase the product, and this was accurately confirmed post-campaign.
* Ancillary data noted include increased general web traffic from non-targeted geographic areas during the campaign period, possibly indicating broader market penetration or incidental traffic increases.
* Other observed metrics included a 10-second average increase in site visit duration compared to the previous period, a 7% increase in social media followers, and a 5% improvement in brand recognition according to post-campaign surveys.

From the complex array of data provided, task is to calculate Precision, Recall, and Accuracy of marketing campaign’s predictive model. Begin by dissecting the narrative to extract and categorize the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) from the described outcomes. Evaluate effectiveness of the model in terms of these metrics, considering broad spectrum of collected data and the viewer behavior analysis.

Q.3 Marks 4

Cluster the following twelve points (with (x, y) representing locations) into four clusters:

(3,10), (3,6), (10,5), (6,8), (8,6), (7,5), (2,3), (5,9), (9,7), (11,4), (3,7), (4,10)

Take initial cluster centers as (3,10), (6,8), (2,3), and (10,5). Provide the final clusters after clustering with K-means clustering. ​

Q.4 Marks 4

Consider the following two-dimensional data points: (1,4), (2,3), (3,6), (4,4), (5,7), (6,5). Find the first principal component using PCA.

Q.5 Marks 4

A university conducted a study to understand the relationship between the number of practice tests completed and the probability of passing the final exam. A logistic regression model was fitted, and the parameters obtained were

β0 = −5 and β1 = 0.8. Calculate the probability that a student will pass the final exam if they complete:

1. 3 practice tests
2. 8 practice tests