DATE: / /
PAGE NO.:

09/08/23

Feature Engineering
Assignment II

Q1 Consider a vector $x = (23, 29, 52, 31, 45, 19, 18, 27)$. Apply feature scaling and find the min-max scaled values and Z-score values.

A (I) Min max scaled values.

$$\text{Formula} = \frac{x - x_{min}}{x_{max} - x_{min}} \left(new_{max} - new\,min \right) + (new\ min).$$

(New max, New min) = (0, 1).

∴

Old Values	New Values
23	0.147
29	0.324
52	1.000
31	0.382
45	0.794
19	0.029
18	0.000
27	0.265

where $x_{min} = 18$; $x_{max} = 52$; $newmin = 0$; $newmax = 1$

(II) Z Score Scaled values.

$$\text{Formula} = \frac{x - \mu}{\sigma}$$

where x = vector value ; μ = mean ; σ = standard deviation.

$\mu = 30.5$; $\sigma = 11.35$

∴

Old Values	New Values
23	−0.661
29	−0.132
52	1.894
31	0.044
45	1.278
19	−1.013
18	−1.101
27	−0.308

Q2 Explain the process of Principle Component Analysis (PCA) in brief.

A (I) Principal Component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning.

(II) PCA generally tries to find the lower-dimensional surface to project high-dimensional data.

(III) Applications of PCA : image processing, movie recommendation system, power allocation optimization.

Steps of PCA.

(I) Getting the dataset

(1) First we need to take the input dataset and divide it into two subparts X and Y, where X is the training dataset and Y is the validation set.

(II) Representing data into a structure

(I) Now we will represent our dataset into a structure. We will represent the two dimensional matrix of independent variable X.

(II) Here, each row corresponds to the data items and the columns correspond to the ~~data items~~ features. The number of columns is the dimension.

(III) Standardizing the data.

(I) The features with higher variance are more important as compared to features with lower variance

(II) If the importance of features is independent of the variance of the feature, then we will divide each data in a column with the standard deviation.

(III) Here, we will name the matrix as Z.

(IV) Calculating the Covariance of Z

(I) We take the matrix Z and transpose it. Then, multiply it by Z.

(II) The output matrix will be the covariance matrix of Z.

(V) Calculating the Eigen Values and Eigen Vectors.

(I) Eigenvectors or the covariance matrix are the directions of the axes with high information.

(II) Also, the coefficients of these eigenvectors are defined as the eigenvalues.

(VI) Sorting the Eigen Vectors

(I) We take all the eigenvalues and sort them in decreasing order.

(II) We sort the eigenvectors accordingly in matrix P of Eigenvalues. The resultant matrix will be named as P*.

(VII) Calculating the new features or principal components.

(I) To do so, we multiply the P* matrix with Z. In the resultant matrix

Z*, each observation is the linear combination of original features.

(ii) Each column of the Z* matrix is independent of one another.

(VII) Remove less or unimportant features from the new dataset.

(i) The new feature set has occurred so we will decide here what to keep and discard.

(ii) We only keep the important or relevant features in the new dataset.

Q3 How to handle the missing values in a dataset that will be used for training the ML model?

A Methods to handle missing values:

(I) Ignore the tuple

When the class label is missing, this technique is used. However, unless the tuple contains numerous attributes with missing values, this approach is not particularly useful.

(II) Fill in the missing values manually

You can replace all the missing values manually. This approach is effective on small data.

(III) Use a global constant to fill in the missing value

You can replace all missing attribute values with a global constant with a label like 'Unknown' or $-\infty$.

(IV) Use a measure of central tendency

Mean or median of the data can be used to fill in the missing value.

(V). Use the most probable value to fill in the missing value
This can be determined using regression, Bayesian classification or decision tree induction.

Q4 Explain the types of wrapper methods for feature selection.

A (I) Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features

(II) It ~~is~~ helps to avoid dimensionality and reduces training time.

(III) Wrapper Methods

① Forward Selection.

(I) Forward selection is an iterative process which begins with an empty set of features.

(II) After each iteration, it keeps adding on a feature and evaluates the performance to check whether it is improving the performance or not.

(III) The process continues until the addition of new features does not improve the performance of the model

② Backward Elimination

(I) Backward Elimination is an iterative approach but it removes the least significant features.

(II) The elimination process continues until removing the features does not improve the performance of the model

③ Exhaustive Feature Selection

(I) It is one of the best selection methods which evaluates each feature set as brute-force.

(II) It means this method tries and makes each possible combination of features and return the best performing feature set.

④ Recursive Feature Elimination:

(I) Recursive feature elimination is a recursive greedy optimization approach, where features are selected by recursively taking smaller subsets.

(II) Now an estimator is trained with each set of features having its importance defined using coef_ attribute.

Q5 Explain Local Binary Pattern (LBP) feature extraction technique with suitable example.

A

(I) Local Binary Pattern (LBP) is a very efficient texture operator which labels the pixels of an image by thresholding the neighbourhood of each ~~image~~ pixel and considers the result as a binary number.

(II) LBP feature ~~&~~ vector:

1	2	2
9	5	6
5	3	1

Threshold →

0	0	0
1		1
1	0	0

Binary : 00010011
Decimal : 19

(I) Divide the examined window into cells (eg 16 x 16 pixels for each cell)

(II) For each pixel in a cell, compare the ~~cell~~ pixels to each of its 8 neighbours. Follow the pixels along a circle i.e. clockwise or counterclockwise

(III) The neighbours considered can be changed by varying the radius of the circle around the pixel R and the quantization of the angular space P.

(IV) If the center pixel's value is greater than the neighbour value, write 0. Else, write 1. This gives can 8 digit binary number.

(V) Compute the histogram over the cell, of the frequency of each 'number' occurring (i.e. each combination of which pixels are smaller and which are greater than the centre). This histogram can be seen as a 256 dimensional feature vector.

(vi) Optionally normalize the histogram.

(vii) Concatenate normalized histograms of all cells. This will give the feature ~~fr~~ vector for the entire window.

(III) The feature vector now can be processed using some machine learning algorithm to classify images. Such classifiers are often used for facial recognition or texture analysis.