PCA

THEORY

1) Principal component analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.
2) Principal component analysis can be broken down into five steps.
   a) Standardize the range of continuous initial variables
   b) Compute the covariance matrix to identify correlations
   c) Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components
   d) Create a feature vector to decide which principal components to keep
   e) Recast the data along the principal components axes
3) Standardization is done by:

$$z = \frac{value - mean}{\sigma} \tag{1}$$

   where σ is standard deviation.

QUIZ