|  |
| --- |
| **Prerequisite**:  Some basics of Math and Statistics before proceeding for PCA:   1. Standard Deviation, Variance, Covariance and Covariance Matrix. 2. Matrix, Eigen Vectors, Eigen Values |
| **Example:** |
| **Principal Component Analysis (PCA)**  Given a set of points, how do we know if they can be compressed like in the previous example? – The answer is to look into the correlation between the points – The tool for doing this is called PCA  An important machine learning method for dimensionality reduction is called Principal Component Analysis It is a method that uses simple matrix operations from linear algebra and statistics to calculate a projection of the original data into the same number or fewer dimensions.  By finding the eigenvalues and eigenvectors of the covariance matrix, we find that the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset. This is the principal component. PCA is a useful statistical technique that has found  application in:  – fields such as face recognition and image compression  – finding patterns in data of high dimension. |
| **PCA Theory in Detail:**  The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.  A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its general objectives are: 1. Dimension Reduction 2. Interpretation.  It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, i.e. by reducing the number of dimensions, without much loss of information.  Dimension reduction offers several benefits such as-   * It compresses the data and thus reduces the storage space requirements. * It reduces the time required for computation since less dimensions require less computation. * It eliminates the redundant features. * It improves the model performance. |
| **Principal Components: Mathematical Details:**  Mathematically, PC’s are particular linear combinations of the p random variables X1, X2,…., Xp. PC’s are solely depend on the covariance matrix (Σ) or the correlation matrix (ρ) of X1, X2 ,…, Xp. It does not require a multivariate normal assumption.  Let the random vector X’ = [X1, X2,…, Xp] have the covariance matrix Σ with the eigenvalues λ1 ≥ λ2 ≥… ≥ λp ≥ 0.  Consider the linear combinations    We obtain,      The principal components are those uncorrelated linear combinations Y1, Y2, … ,Yp whose variances are as large as possible.  First Principal Component: Y1 = a1’ X that maximizes V(a1’ X) subject to a1’a1  = 1.  Second Principal Component: Y2 = a2’ X that maximizes V(a2’ X) subject to a2’a2  = 1 and  Cov(a1’ X, a2’ X) = 0.  .  .  .  ith Principal Component: Yi = ai’ X that maximizes V(ai’ X) subject to ai’ai  = 1 and  Cov(ai’ X, ak’ X) = 0.      The proportion of total variance explained due to k­th principal component is |
| **Algorithm behind PCA**  STEP 1: STANDARDIZATION  The aim of this step is to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis.  More specifically, the reason why it is critical to perform standardization prior to PCA, is that the latter is quite sensitive regarding the variances of the initial variables. That is, if there are large differences between the ranges of initial variables, those variables with larger ranges will dominate over those with small ranges (For example, a variable that ranges between 0 and 100 will dominate over a variable that ranges between 0 and 1), which will lead to biased results. So, transforming the data to comparable scales can prevent this problem.  Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable.  Once the standardization is done, all the variables will be transformed to the same scale.  STEP 2: COVARIANCE MATRIX COMPUTATION  The aim of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them. Because sometimes, variables are highly correlated in such a way that they contain redundant information. So, in order to identify these correlations, we compute the covariance matrix.  The covariance matrix is a *p* × *p* symmetric matrix (where *p*is the number of dimensions) that has as entries the covariance associated with all possible pairs of the initial variables. For example, for a 3-dimensional data set with 3 variables *x*, *y*, and *z*, the covariance matrix is a 3×3 matrix of this from:  Covariance Matrix for 3-Dimensional Data  Covariance Matrix for 3-Dimensional Data  Since the covariance of a variable with itself is its variance (Cov(a,a)=Var(a)), in the main diagonal (Top left to bottom right) we actually have the variances of each initial variable. And since the covariance is commutative (Cov(a,b)=Cov(b,a)), the entries of the covariance matrix are symmetric with respect to the main diagonal, which means that the upper and the lower triangular portions are equal. STEP 3: COMPUTE THE EIGENVECTORS AND EIGENVALUES OF THE COVARIANCE MATRIX TO IDENTIFY THE PRINCIPAL COMPONENTS Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the principal components of the data. Before getting to the explanation of these concepts, let’s first understand what do we mean by principal components.  Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables. These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components. So, the idea is 10-dimensional data gives you 10 principal components, but PCA tries to put maximum possible information in the first component, then maximum remaining information in the second and so on, until having something like shown in the scree plot below.    Organizing information in principal components this way, will allow you to reduce dimensionality without losing much information, and this by discarding the components with low information and considering the remaining components as your new variables.  An important thing to realize here is that, the principal components are less interpretable and don’t have any real meaning since they are constructed as linear combinations of the initial variables.  Geometrically speaking, principal components represent the directions of the data that explain a **maximal amount of variance**, that is to say, the lines that capture most information of the data. The relationship between variance and information here, is that, the larger the variance carried by a line, the larger the dispersion of the data points along it, and the larger the dispersion along a line, the more the information it has. To put all this simply, just think of principal components as new axes that provide the best angle to see and evaluate the data, so that the differences between the observations are better visible.  Example:  let’s suppose that our data set is 2-dimensional with 2 variables *x,y*and that the eigenvectors and eigenvalues of the covariance matrix are as follows:  Principal Component Analysis Example  If we rank the eigenvalues in descending order, we get λ1>λ2, which means that the eigenvector that corresponds to the first principal component (PC1) is *v1*and the one that corresponds to the second component (PC2) is*v2.*  After having the principal components, to compute the percentage of variance (information) accounted for by each component, we divide the eigenvalue of each component by the sum of eigenvalues. If we apply this on the example above, we find that PC1 and PC2 carry respectively 96% and 4% of the variance of the data. STEP 4: FEATURE VECTOR As we saw in the previous step, computing the eigenvectors and ordering them by their eigenvalues in descending order, allow us to find the principal components in order of significance. In this step, what we do is, to choose whether to keep all these components or discard those of lesser significance (of low eigenvalues), and form with the remaining ones a matrix of vectors that we call Feature vector.  So, the feature vector is simply a matrix that has as columns the eigenvectors of the components that we decide to keep. This makes it the first step towards dimensionality reduction, because if we choose to keep only p eigenvectors (components) out of n, the final data set will have only p dimensions.  Example:  Continuing with the example from the previous step, we can either form a feature vector with both of the eigenvectors *v*1 and *v*2:  Principal Component Analysis eigen vectors  Or discard the eigenvector *v*2, which is the one of lesser significance, and form a feature vector with *v*1 only:  Principal Component Analysis eigen vectors 2  Discarding the eigenvector *v2*will reduce dimensionality by 1, and will consequently cause a loss of information in the final data set. But given that *v*2 was carrying only 4% of the information, the loss will be therefore not important and we will still have 96% of the information that is carried by *v*1. LAST STEP: RECAST THE DATA ALONG THE PRINCIPAL COMPONENTS AXES In the previous steps, apart from standardization, you do not make any changes on the data, you just select the principal components and form the feature vector, but the input data set remains always in terms of the original axes (i.e, in terms of the initial variables).  In this step, which is the last one, the aim is to use the feature vector formed using the eigenvectors of the covariance matrix, to reorient the data from the original axes to the ones represented by the principal components (hence the name Principal Components Analysis). This can be done by multiplying the transpose of the original data set by the transpose of the feature vector.  Principal Component Analysis feature vector |
| **Working Rule:**   1. Find out average of rows 2. Find out centered matrix A 3. find out Covariance Matrix S= 4. Find out eigen values of S. 5. Find out eigen vector of S corresponding to larger eigen value of S. 6. Normalize it. |
| **Classroom Problems:**  Q.01 Construct the sample covariance matrix S for the data given below:    Find the eigenvector of S that points the most significant direction of the data  Solution:    Q. 02 Construct the sample covariance matrix S for the data given below:    Find the eigenvector of S that points the most significant direction of the data  Solution:  E:\NMIMS\Maths\Linear Algebra\Unit V\PCA2.jpeg    Q. 03 Construct the sample covariance matrix S for the data given below:    Find the eigenvector of S that points the most significant direction of the data  Solution    Q.04 <https://www.gatevidyalay.com/tag/principal-component-analysis-example-ppt/>  Q.05        Q.06 |