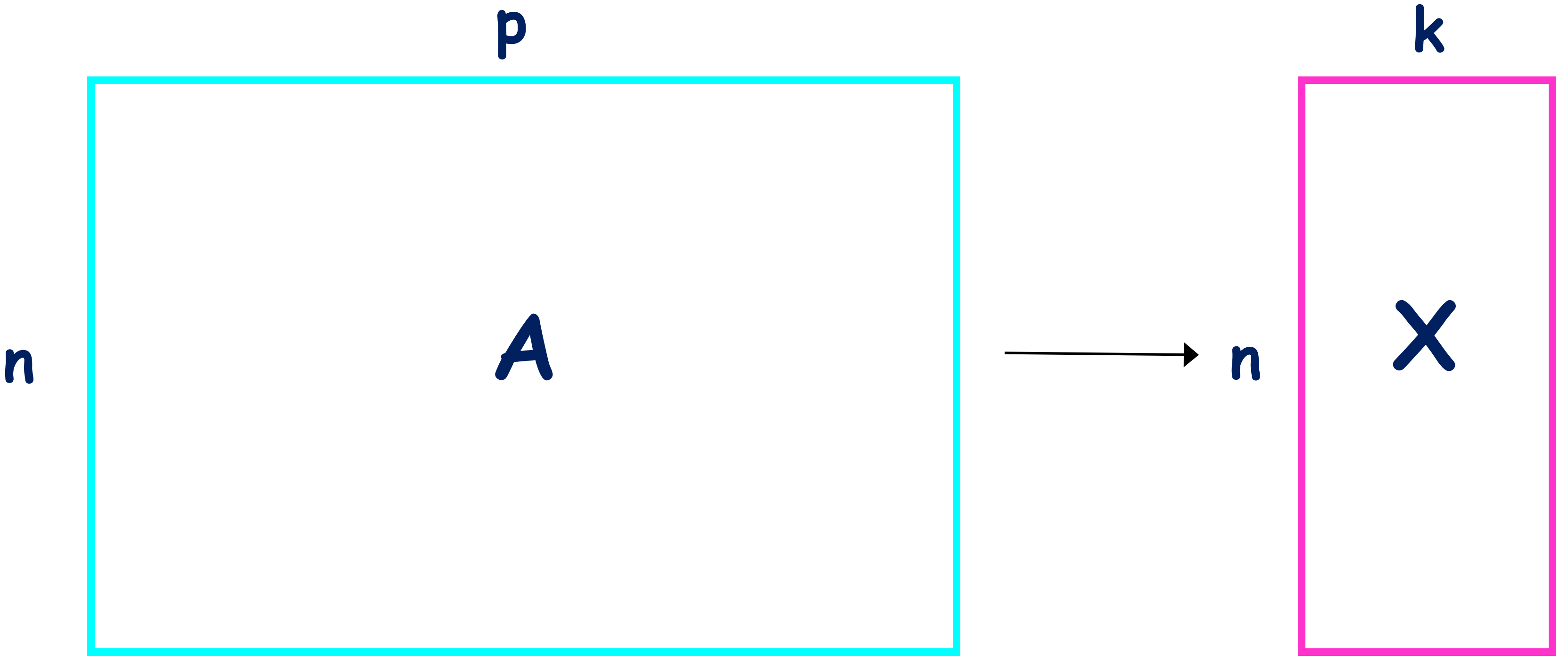
LINEAR ALGEBRA FOR AI/ML

JIAUL PAIK

Principal Component Analysis (PCA)

Data Reduction

- Summarization of data with many (p) variables by a smaller set of (k) derived (latent, composite) variables.



Data Presentation: Key questions?

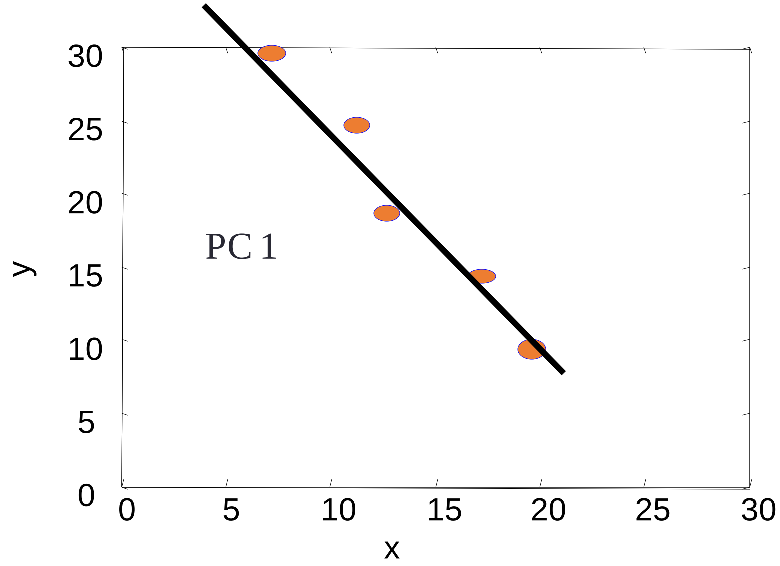
- Do we need a all n-dimension space to view data?
- Better presentation (new axes) than original axes?
- How to find the ‘best’ low dimension space that conveys maximum useful information?
- One answer: Find “Principal Components”

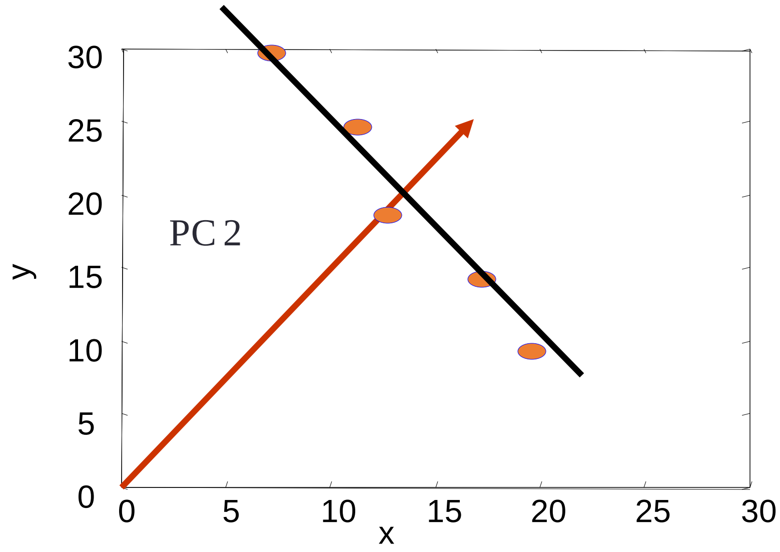
Principal Components

- All principal components (PCs) start at the origin

- First PC is direction of maximum variance

- Subsequent PCs are orthogonal to 1st PC and describe maximum residual variance



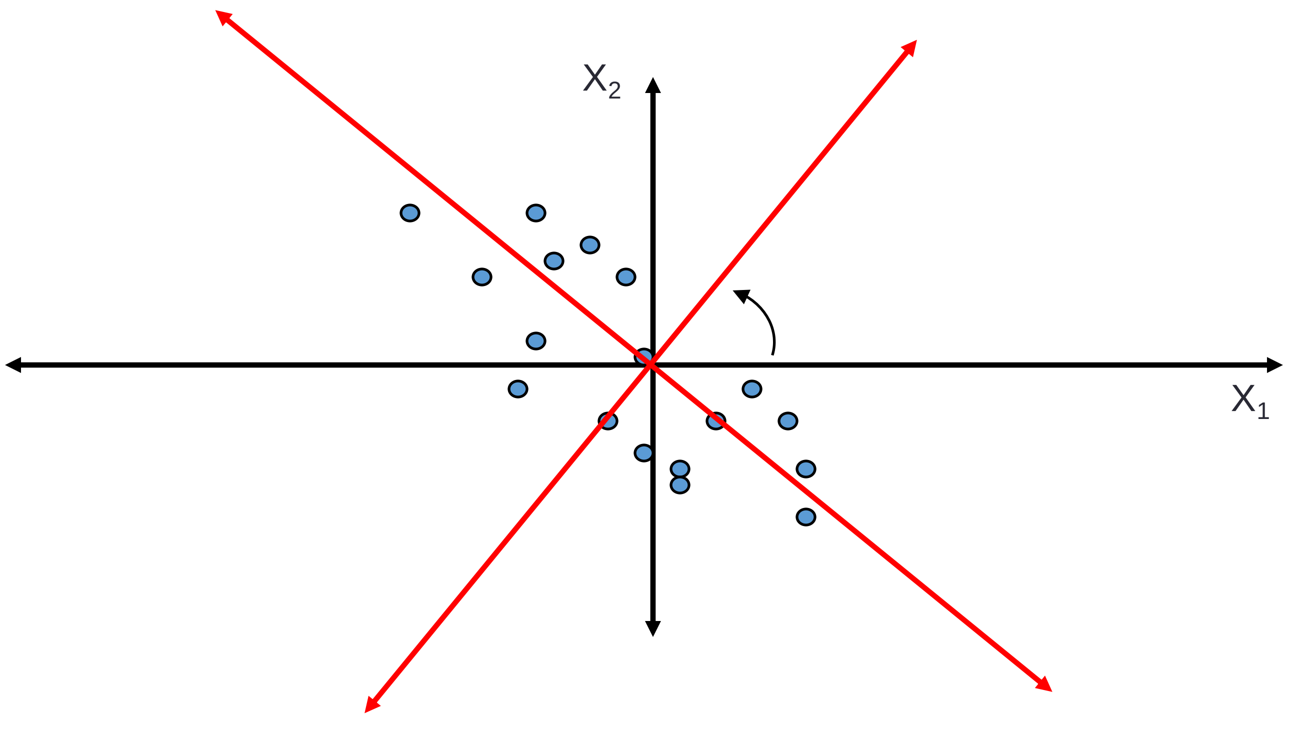


The Goal

Summarize the underlying variance-covariance structure of a large set of variables through a few linear combinations of these variables.

Trick: Rotate Coordinate Axes

- Suppose we have p features $x_1, \ldots, x_p$.
- Our goal is to develop a new set of p axes (linear combinations of the original p axes) in the directions of greatest variability:
- This is accomplished by rotating the axes.



Principal Component Analysis (PCA)

- Takes a data matrix of *n* objects by *p* variables
- Variables may be correlated, and summarizes it by uncorrelated axes
    - That are linear combinations of the original *p* variables
    - The first *k* components display as much as possible of the variation among objects.

Geometric Rationale of PCA

- Objects are points in a multidimensional space with an axis for each of the *p* variables

- The centroid of the points is defined by the mean of each variable

- The variance of each variable is the average squared deviation of its *n* values around the mean of that variable.

$$V_i = \frac{1}{n-1}\sum_{m=1}^{n}\left(X_{im} - \bar{X}_i\right)^2$$

Geometric Rationale of PCA

- Covariance: degree to which the variables are linearly correlated to each other

$$C_{ij} = \frac{1}{n}\sum_{m=1}^{n}\left(X_{im} - \overline{X}_{i}\right)\left(X_{jm} - \overline{X}_{j}\right)$$

Covariance of variables *i* and *j*

Sum over all *n* objects

Value of variable *i* in object *m*

Mean of variable *i*

Value of variable *j* in object *m*

Mean of variable *j*

Generalization to *p*-dimensions

- The algebra for finding principal axes readily generalizes to *p* variables
  - PC 1 is the direction of maximum variance in the *p*-dimension
  - PC 2 is in the direction of the next highest variance, subject to the constraint that it has zero covariance with PC 1.
  - PC 3 is in the direction of the next highest variance, subject to the constraint that it has zero covariance with both PC 1 and PC 2
  - and so on... up to PC p

Covariance: An Important Issue

- Covariance among variables makes sense if they are measured in the same units
  - variables with high variances will dominate the principal components
- It is avoided by standardizing each variable to unit variance and zero mean.

$$X'_{im} = \frac{\left(X_{im} - \overline{X}_i\right)}{\mathrm{SD}_i}$$

Mean variable *i*

Standard deviation of variable *i*

Covariance *vs* Correlation

- Covariances between the standardized variables are correlations
- After standardization, each variable has a variance of 1
- Correlations can be also calculated from the variances and covariances:

Correlation co-eff.

$$r_{ij} = \frac{C_{ij}}{\sqrt{V_i V_j}}$$

Covariance of variables *i* and *j*

Variance of variables

The Algebra of PCA

- First step is to calculate the matrix of variances and covariances (or correlations) among every pair of the *p* variables

- Square, symmetric matrix

- Diagonals are the variances, off-diagonals are the co-variances.

	X_1	X_2
X_1	6.6707	3.4170
X_2	3.4170	6.2384

Variance-covariance Matrix

The Algebra of PCA

- In matrix notation, this is computed as (assuming you have zero mean matrix)

$$\boldsymbol{S} = \boldsymbol{X}^{T}\boldsymbol{X}$$

where X is the *n* x *p* data matrix, with each variable centered

	X_1	X_2
X_1	6.6707	3.4170
X_2	3.4170	6.2384

Variance-covariance Matrix

The Algebra of PCA

- Sum of the diagonals of the variance-covariance matrix is called the trace

- It represents the total variance in the data

	X_1	X_2
X_1	6.6707	3.4170
X_2	3.4170	6.2384

Trace = 12.9091

The Algebra of PCA

- Finding the principal axes involves Eigen analysis of the covariance matrix (S)

- The eigenvalues of S are solutions (λ) to the characteristic equation

$$\left|\mathbf{S} - \lambda\mathbf{I}\right| = 0$$

The Algebra of PCA

- The eigenvalues, λ_1, λ_2, ... λ_p are the variances of the coordinates on each principal component axis

- The sum of all *p* eigenvalues equals the trace of S (the sum of the variances of the original variables).

	X_1	X_2
X_1	6.6707	3.4170
X_2	3.4170	6.2384

λ_1 = 9.8783
λ_2 = 3.0308

Note: $\lambda_1 + \lambda_2$ =12.9091

Trace = 12.9091

The Algebra of PCA

- Each eigenvector consists of *p* values which represent the "contribution" of each variable to the principal component axis

- Eigenvectors are uncorrelated (orthogonal)
  - their dot-products are zero.

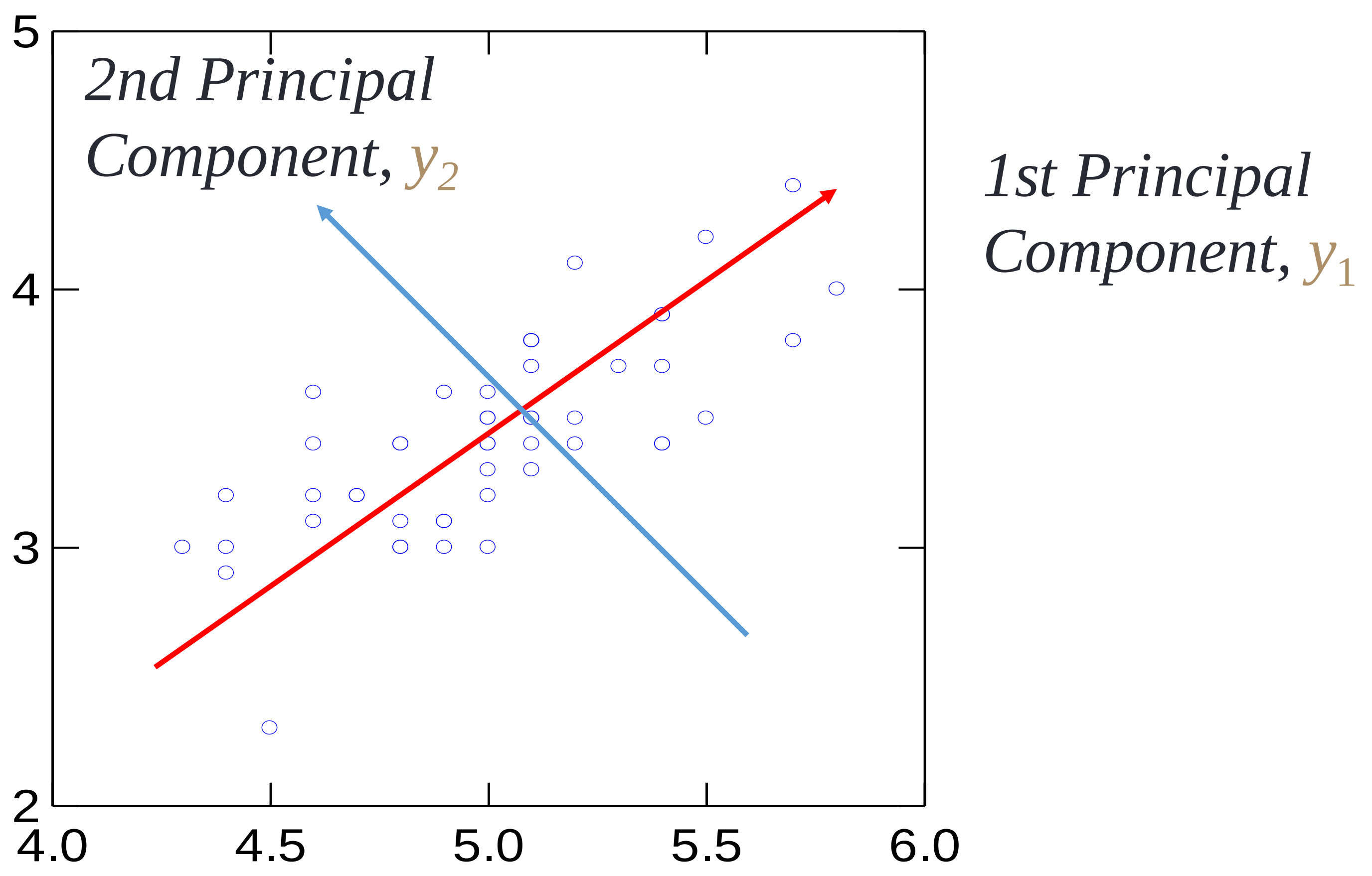
Eigenvectors

	u_1	u_2
X_1	0.7291	-0.6844
X_2	0.6844	0.7291

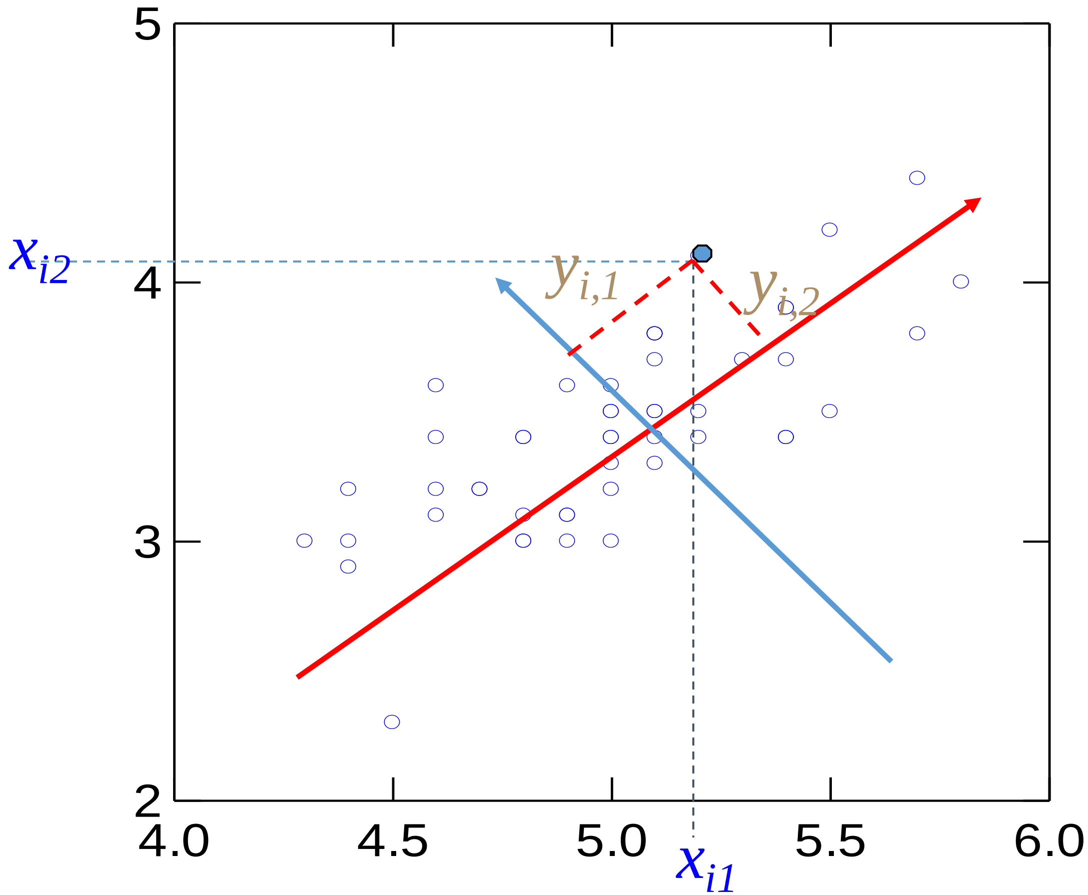
$0.7291*(-0.6844) + 0.6844*0.7291 = 0$

Product of eigenvectors

Projecting Data to Lower Dimension



Projecting Data to Lower Dimension

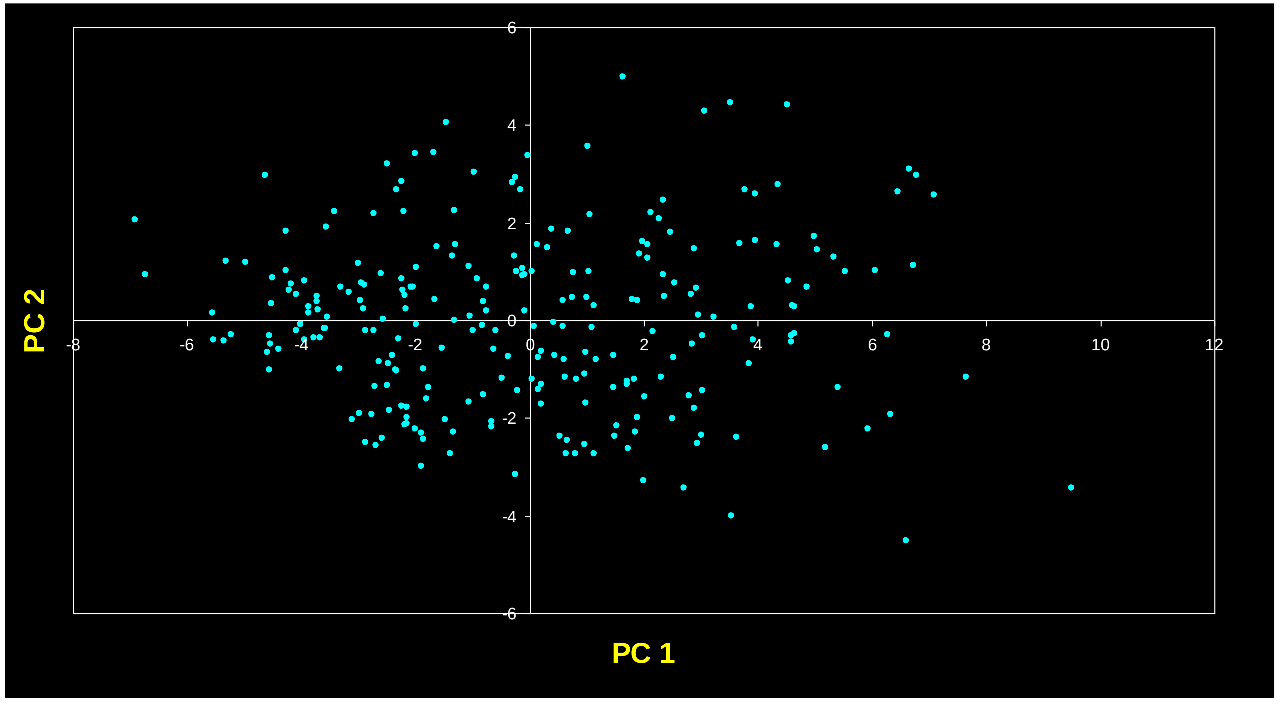


Contribution of Principal Component

$\lambda_1 = 9.8783$ $\lambda_2 = 3.0308$ Trace = 12.9091

PC 1 displays

9.8783/12.9091 = 76.5% of the total variance



PCA: Running Example

PCA Example –STEP 1

$$(v - v_{mean})/sd(v)$$

Raw Data

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1.0	1.1
1.5	1.6
1.1	0.9

Zero Mean Data

x	y
0.69	0.49
-1.31	-1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

PCA Example –STEP 2

- Calculate the covariance matrix

$$\text{cov} = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

PCA Example –STEP 3

- Calculate the eigenvectors and eigenvalues of the covariance matrix

$$\text{eigenvalues} = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$$\text{eigenvectors} = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

PCA Example –STEP 4

- Reduce dimensionality and form *feature vector*

  the eigenvector with the *highest* eigenvalue is the *principle component* of the data set.

- Once eigenvectors are found from the covariance matrix, the next step is to
  - order them by eigenvalue, highest to lowest.
    - this gives you the components in order of significance.

PCA Example –STEP 4

- Feature Vector: Feature Vector = (eig_1 eig_2 eig_3 … eig_n)

We can either form a feature vector with both of the eigenvectors:

$$\begin{pmatrix} -.677873399 & -.735178656 \\ -.735178656 & .677873399 \end{pmatrix}$$

or, we can choose to leave out the smaller, less significant component and only have a single column:

$$\begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix}$$

PCA Example –STEP 5

Final Data

x	y
-.827970186	-.175115307
1.77758033	.142857227
-.992197494	.384374989
-.274210416	.130417207
-1.67580142	-.209498461
-.912949103	.175282444
.0991094375	-.349824698
1.14457216	.0464172582
.438046137	.0177646297
1.22382056	-.162675287

How many axes are needed?

- Does the $(k+1)^{th}$ principal axis represent more variance than would be expected by chance?

- A common "rule of thumb" when PCA is based on correlations is that axes with eigenvalues > 1 are worth interpreting

What are the assumptions of PCA?

- Assumes relationships among variables are linear
  - Points in *p*-dimensional space has linear dimensions that can be effectively summarized by the principal axes

- If the structure in the data is non-linear, the principal axes will not be an efficient and informative summary of the data.

THANK YOU Students....