Principal Component Analysis

Objectives

After completing this module, you should be able to:

Understand Covariance, Eigen Values & Eigen Vectors

Make use of Dimensionality Reduction Techniques

Apply Principal Component Analysis

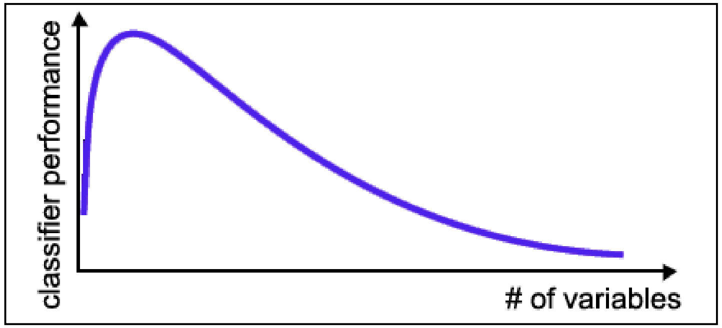
Decide when to use PCA

Curse Of Dimensionality

The required number of samples (to achieve the same accuracy) grows exponentionally with the number of variables!

In practice: number of training examples is fixed!

=> the classifier's performance usually will degrade for a large number of features!



Covariance

Variance and Covariance are a measure of the "spread" of a set of points around their center of mass (mean)

Variance – measure of the deviation from the mean for points in one dimension e.g. heights

Covariance as a measure of how much each of the dimensions vary from the mean with respect to each other.

Covariance is measured between 2 dimensions to see if there is a relationship between the 2 dimensions e.g. number of hours studied & marks obtained.

The covariance between one dimension and itself is the variance

4

Covariance

$$cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})(y-\bar{y})$$

So, if you had a 3-dimensional data set (x,y,z), then you could measure the covariance between the x and y dimensions, the y and z dimensions, and the x and z dimensions.

Measuring the covariance between x and x , or y and y , or z and z would give you the variance of the x , y and z dimensions respectively.

Covariance

Representing Covariance between dimensions as a matrix e.g. for 3 dimensions:

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix}$$

Diagonal is the variances of x, y and z

cov(x,y) = cov(y,x) hence matrix is symmetrical about the diagonal

N-dimensional data will result in NxN covariance matrix

6

Eigen Values and Eigen Vectors

In linear algebra, an eigenvector or characteristic vector of a linear transformation is a non-zero vector whose direction does not change when that linear transformation is applied to it.

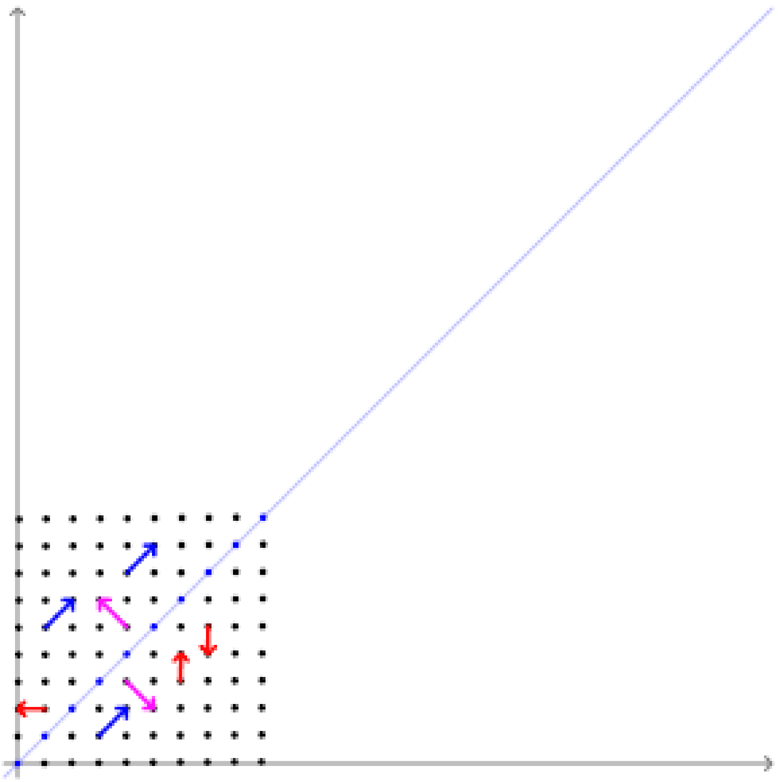
More formally, if T is a linear transformation from a vector space V over a field F into itself, then v is an eigenvector of T.

$T(v)= \lambda V$

where λ is a scalar in the field F, known as the eigenvalue, characteristic value, or characteristic root associated with the eigenvector v.

7

Eigen Vectors



8

Example: how to find eigenvalues and eigenvectors

let the covariance matrix is : $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$

First Step: compute eigenvalues. To do this, we find the values of which satisfy the characteristic equation of the matrix A, namely those values of for which

$$\det(A - \lambda I) = 0,$$

1. $\lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

2. $A - \lambda I = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$

3. $\det \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} = (7-\lambda)(-1-\lambda)-(3)(3) = -7-7\lambda+\lambda+\lambda^2-9 = \boxed{\lambda^2-6\lambda-16}$

Let this equation equal to zero and solve it

4. $(\lambda-8)(\lambda+2)=0 \longrightarrow \underbrace{\boxed{\lambda_1=8} \quad \boxed{\lambda_2=-2}}_{\text{EigenValues}}$

9

Second step: find corresponding eigenvectors.

1. For each eigenvalue λ, we have $(A - \lambda I)V = 0$, where x is the eigenvector associated with eigenvalue λ.

2. Find x by Gaussian elimination. That is, convert the augmented matrix

$$(A - \lambda I \vdots 0)$$

to row echelon form, and solve the resulting linear system by back substitution.

3. We find the eigenvectors associated with each of the eigenvalues

10

When λ1=8

$$A-\lambda I = \begin{bmatrix} 7-8 & 3 \\ 3 & -1-8 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}$$

$$(A - \lambda I)V = 0 \longrightarrow \begin{bmatrix} -1 & 3 & \vdots & v_1 \\ 3 & -9 & \vdots & v_2 \end{bmatrix} = 0$$

$$(\mathbf{-1 \times v_1}) + (\mathbf{3 \times v_2}) = \mathbf{0}$$

$$(\mathbf{3 \times v_1}) + (\mathbf{-9 \times v_2}) = \mathbf{0}$$

$$+\quad \begin{aligned} -1v_1 + 3v_2 &= 0 \quad \text{..........(1} \\ 3v_1 - 9v_2 &= 0 \quad \text{..........(2} \\ \hline 2v_1 - 6v_2 &= 0 \quad \text{..........(3} \end{aligned}$$

Lets $s = \sqrt{(2)^2 + (-6)^2} = 6.3246$

$v_1 = 2/6.3246 = 0.3162$

$v_2 = -6/6.3246 = -0.9487$

11

When λ1=-2

$$A-\lambda I=\begin{bmatrix} 7+2 & 3 \\ 3 & -1+2 \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$$

$$(A-\lambda I)V = 0 \Longrightarrow \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \vdots \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$\mathbf{(9 \times v_1) + (3 \times v_2) = 0}$$

$$\mathbf{(3 \times v_1) + (1 \times v_2) = 0}$$

$$9v_1 + 3v_2 = 0 \quad \text{..........(1}$$

$$+ \quad 3v_1 + 1v_2 = 0 \quad \text{..........(2}$$

$$6v_1 + 2v_2 = 0 \quad \text{..........(3}$$

Lets $s = \sqrt{(2)^2 + (6)^2} = 6.3246$

v_1 = -6/6.3246= **-0.9487**

v_2 = -2/6.3246= **-0.3162**

12

Principal Component Analysis

Principal component analysis (PCA) is a way to reduce data dimensionality

PCA projects high dimensional data to a lower dimension

PCA projects the data in the least square sense– it captures big (principal) variability in the data and ignores small variability

13

PCA algorithm I (sample covariance matrix)

Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, compute covariance matrix $\mathbf{\Sigma}$

$$\boxed{\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T}$$

where

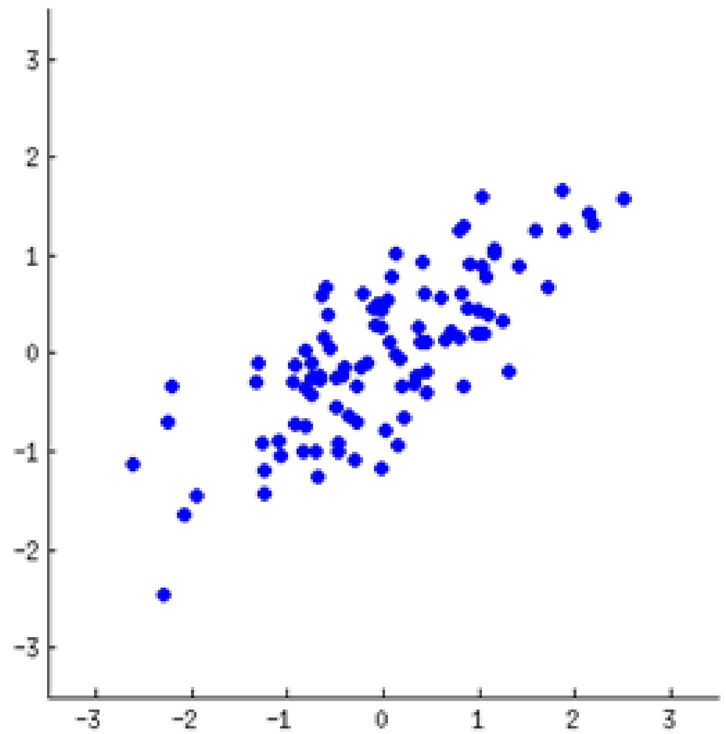
$$\boxed{\overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i}$$

PCA basis vectors = the eigenvectors of $\mathbf{\Sigma}$

Larger eigenvalue $\Rightarrow$ more important eigenvectors

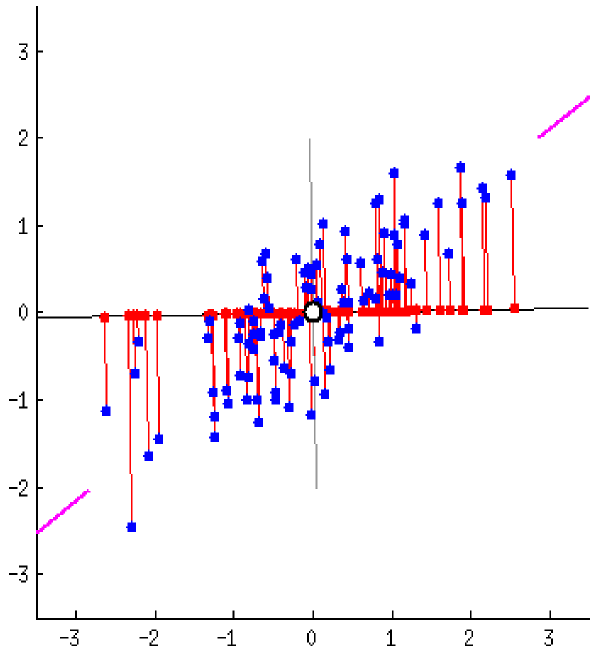
14

How does PCA Work



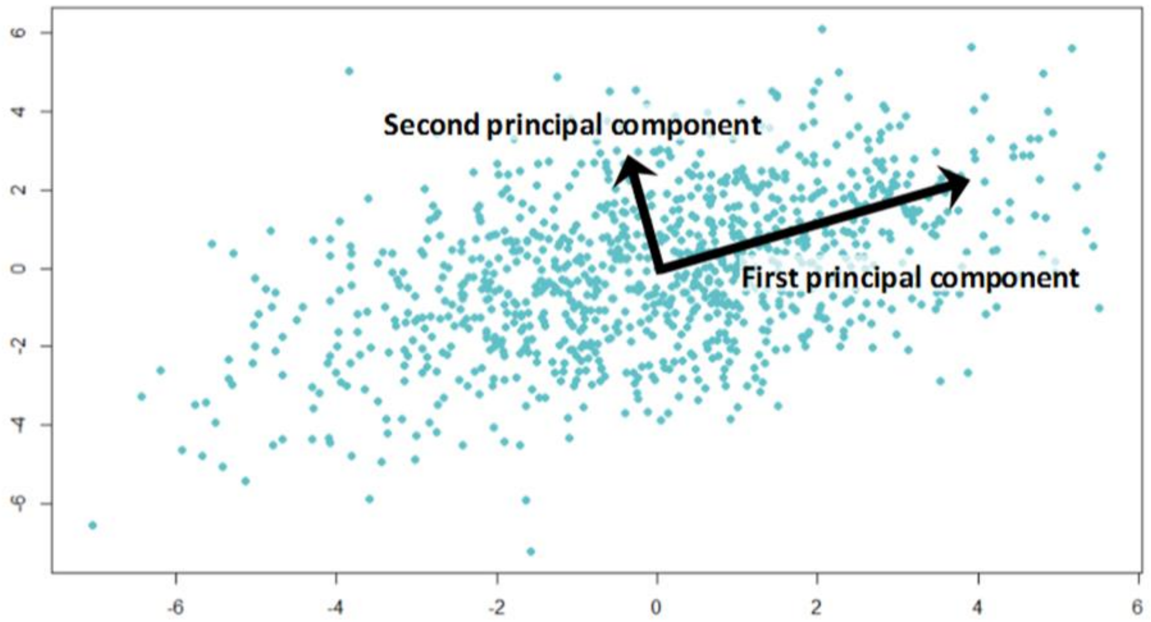
15

How does PCA Work



16

Principal Component



PCA

Principal component analysis (PCA) is a way to reduce data dimensionality

PCA projects high dimensional data to a lower dimension

PCA projects the data in the least square sense– it captures big (principal) variability in the data and ignores small variability

Given data $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, compute covariance matrix $\boldsymbol{\Sigma}$

PCA basis vectors = the eigenvectors of $\boldsymbol{\Sigma}$

Larger eigenvalue $\Rightarrow$ more important eigenvectors

18
18

PCA Example

1. Given original data set S = {$\mathbf{x}^1, \ldots, \mathbf{x}^k$}, produce new set by subtracting the mean of attribute A_i from each x_i.

Data =

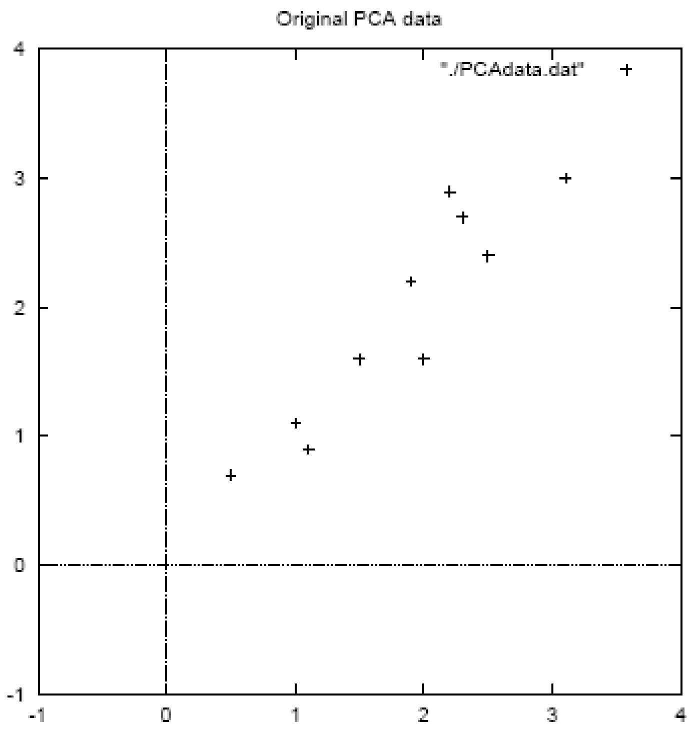
x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9
Mean: 1.81	**1.91**

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01
Mean: 0	**0**

19

PCA Example



20

PCA Example

2. Calculate the covariance matrix:

$$
cov = \begin{array}{c c} & \begin{array}{cc} \mathbf{x} & \mathbf{y} \end{array} \\ \begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array} & \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix} \end{array}
$$

3. Calc [illegible] ance matrix:

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

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

21

PCA Example

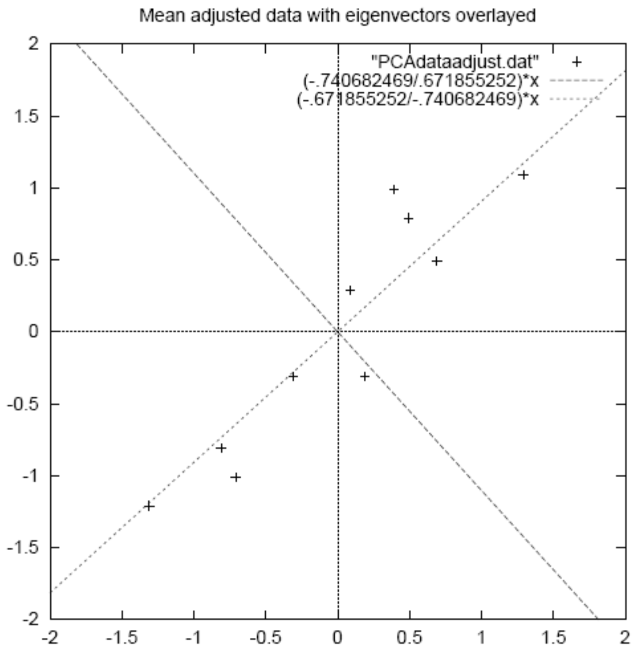


Figure 3.2: A plot of the normalised data (mean subtracted) with the eigenvectors of the covariance matrix overlayed on top.

- Eigenvector with largest eigenvalue traces linear pattern in data

- 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.

22

PCA Example

4. Order eigenvectors by eigenvalue, highest to lowest.

$$\mathbf{v}_1 = \begin{pmatrix} -.677873399 \\ -.735178956 \end{pmatrix} \quad \lambda = 1.28402771$$

$$\mathbf{v}_2 = \begin{pmatrix} -.735178956 \\ .677873399 \end{pmatrix} \quad \lambda = .0490833989$$

In general, you get *n* components. To reduce dimensionality to *p*, ignore $n-p$ components at the bottom of the list.

23

PCA Example

Construct new feature vector.

Feature vector = ($\mathbf{v}_1$, $\mathbf{v}_2$, ...$\mathbf{v}_p$)

$$FeatureVector1 = \begin{pmatrix} -.677873399 & -.735178956 \\ -.735178956 & .677873399 \end{pmatrix}$$

or reduced dimension feature vector :

$$FeatureVector2 = \begin{pmatrix} -.677873399 \\ -.735178956 \end{pmatrix}$$

24

PCA Example

5. Derive the new data set.

$$RowFeatureVector1 = \begin{pmatrix} -.677873399 & -.735178956 \\ -.735178956 & .677873399 \end{pmatrix}$$

TransformedData = RowFeatureVector × RowDataAdjust

$$RowFeatureVector2 = \begin{pmatrix} -.677873399 & -.735178956 \end{pmatrix}$$

$$RowDataAdjust = \begin{pmatrix} .69 & -1.31 & .39 & .09 & 1.29 & .49 & .19 & -.81 & -.31 & -.71 \\ .49 & -1.21 & .99 & .29 & 1.09 & .79 & -.31 & -.81 & -.31 & -1.01 \end{pmatrix}$$

This gives original data in terms of chosen components (eigenvectors)—that is, along these axes.

PCA Example

Transformed 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

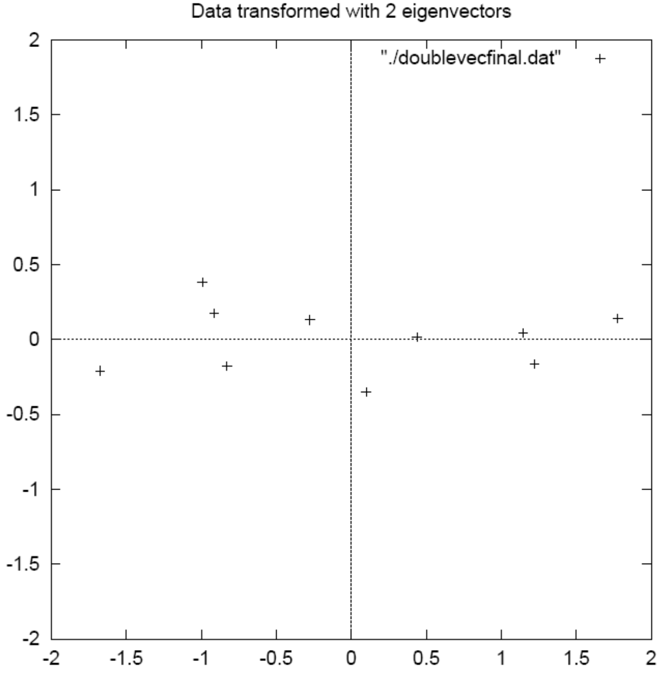


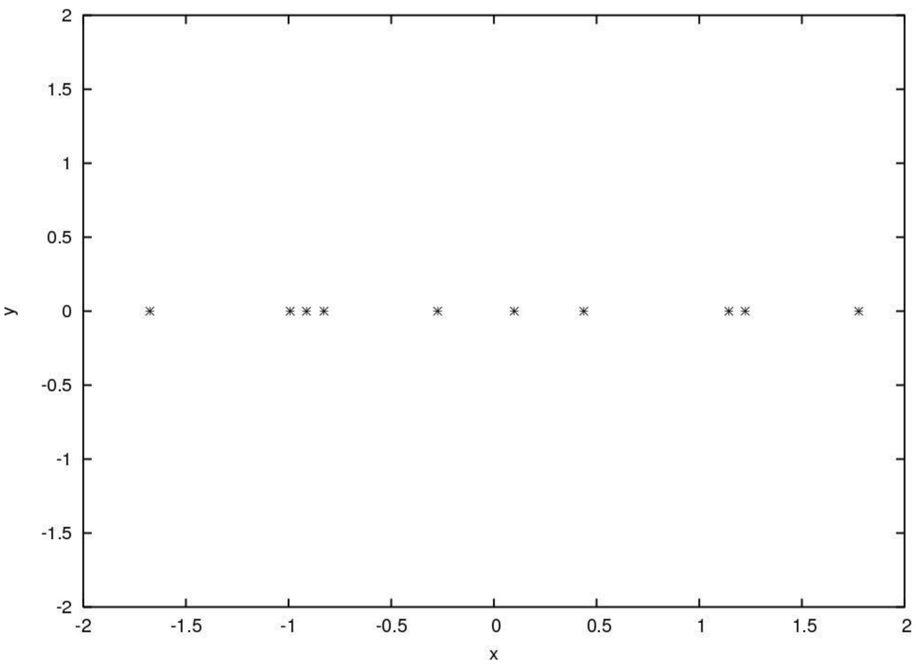
Figure 3.3: The table of data by applying the PCA analysis using both eigenvectors, and a plot of the new data points.

26

PCA Example - Result

Transformed Data (Single eigenvector)

x
-.827970186
1.77758033
-.992197494
-.274210416
-1.67580142
-.912949103
.0991094375
1.14457216
.438046137
1.22382056



27

Reconstructing the original data

We did:

$TransformedData = RowFeatureVector \times RowDataAdjust$

so we can do

$RowDataAdjust = RowFeatureVector^{-1} \times TransformedData$

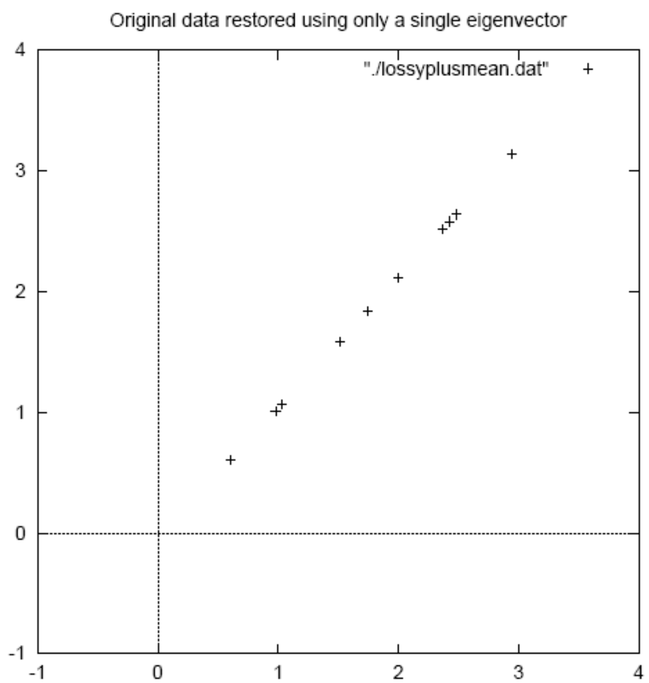
$= RowFeatureVector^{T} \times TransformedData$

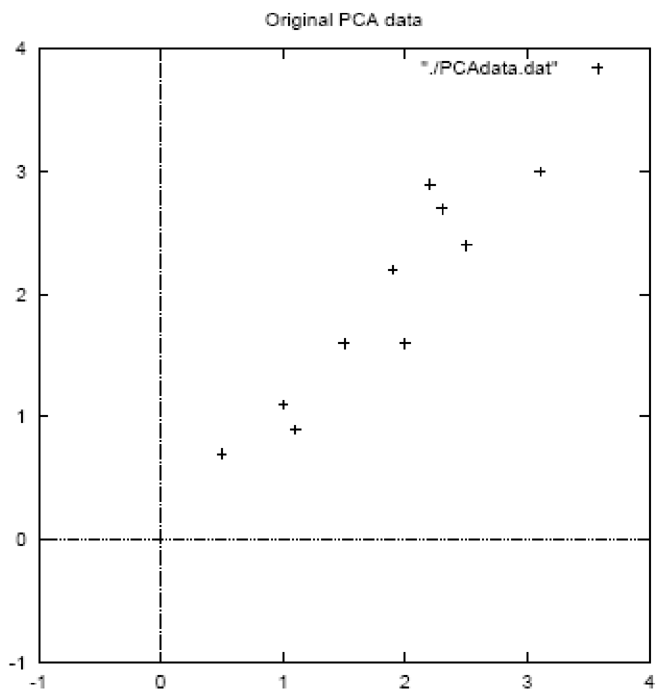
and

$RowDataOriginal = RowDataAdjust + OriginalMean$

28

Reconstructed Data





29

Applications of PCA

- Data Visualization
- Data Compression - Reduce memory/space needed to store data
- Data Compression - Speedup your learning algorithm
- Dimensionality reduction

30

Summary

Covariance

Eigen Values & Eigen Vectors

Dimensionality Reduction

Principal Component Analysis

Mathematical Modelling – PCA Algorithm 1

PCA Algorithm 1 – Example

Mathematical Modelling - PCA Algorithm II

Applications of PCA

Thank you