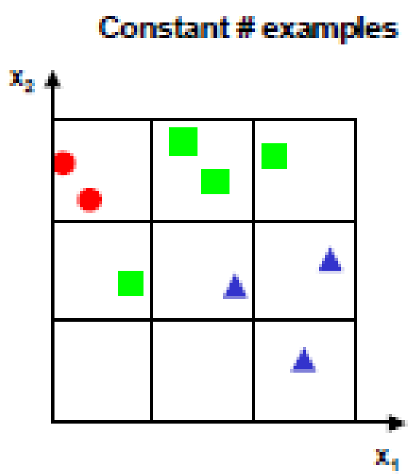
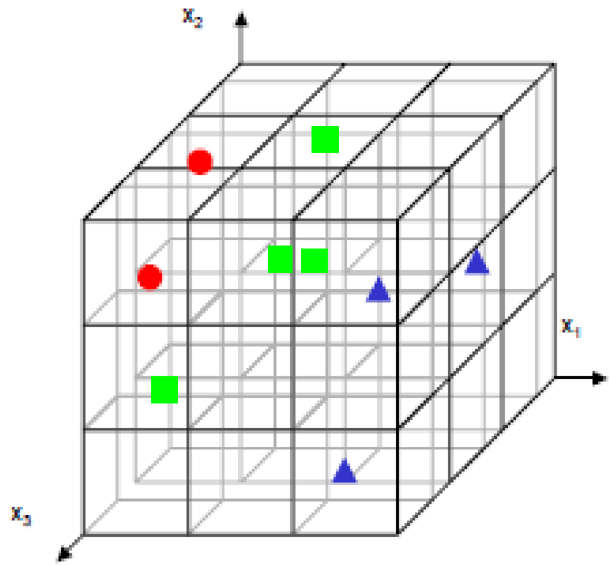
Tutorial on Principal Components Analysis

Daniele Cerra, DLR

Curse of Dimensionality

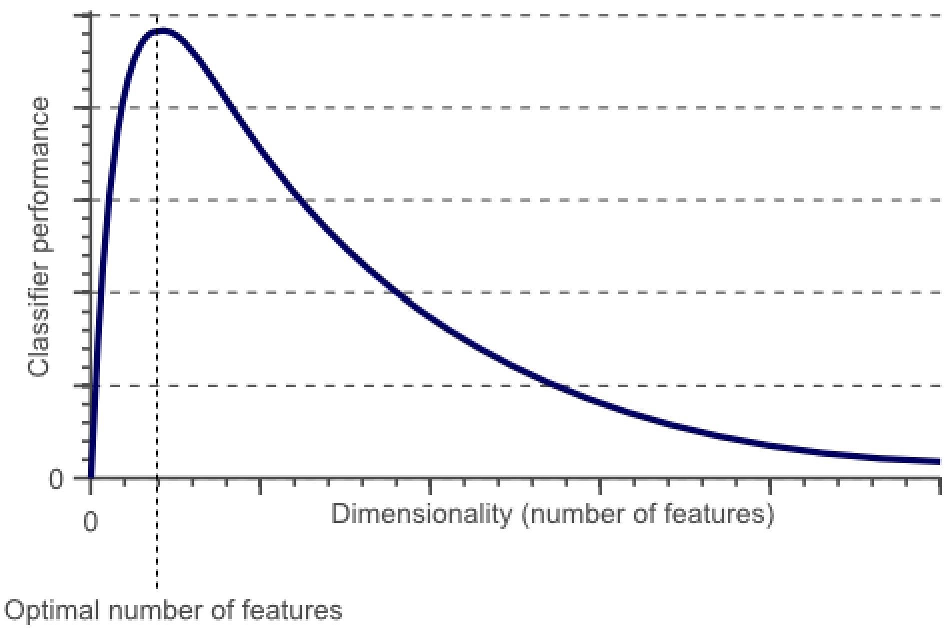
– Classification problem: 3 classes



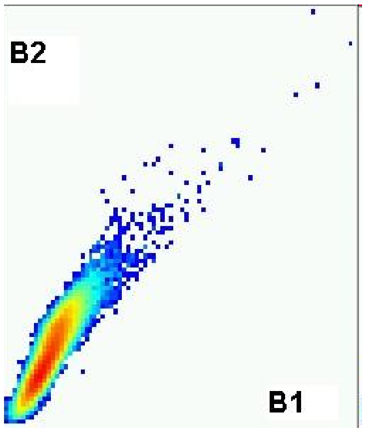


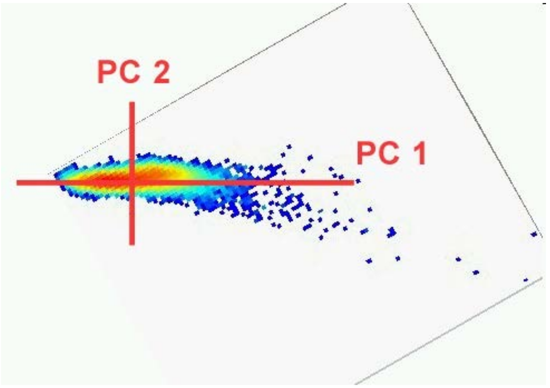
Curse of Dimensionality

– Classification problem



Really the whole spectrum? Second part

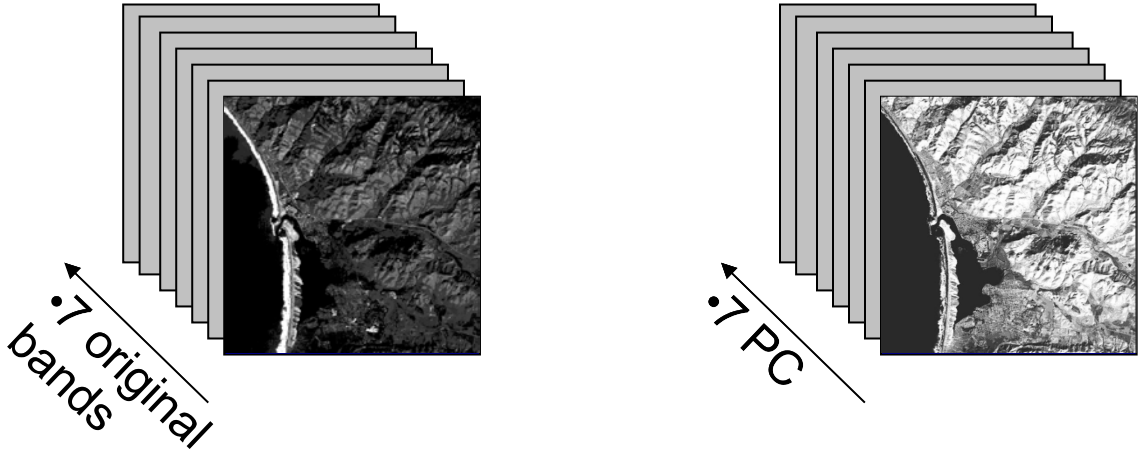




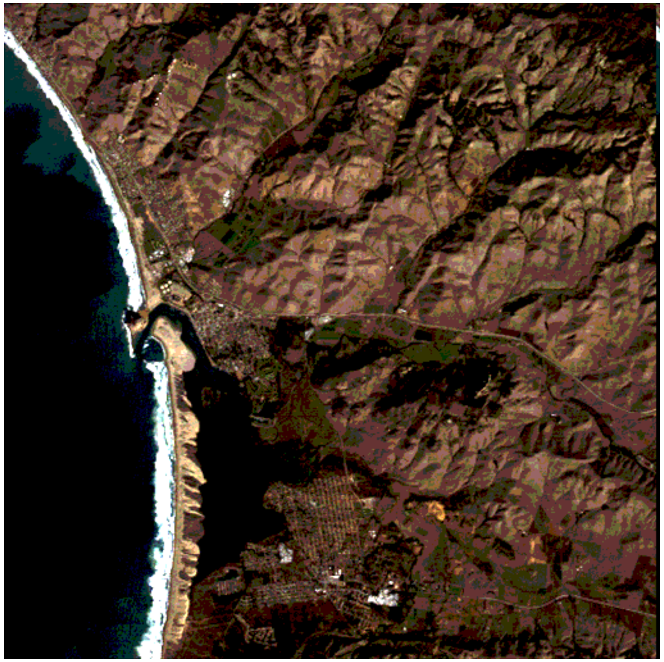
- Correlation between bands
  - Information redundancy

- We "change" the bands to have independent information in the bands
- By rotating the feature space
- Then we select new "bands" with high variance only

Principal Components Analysis



A typical Landsat Image



•Morro Bay, California, USA

•RGB Combination

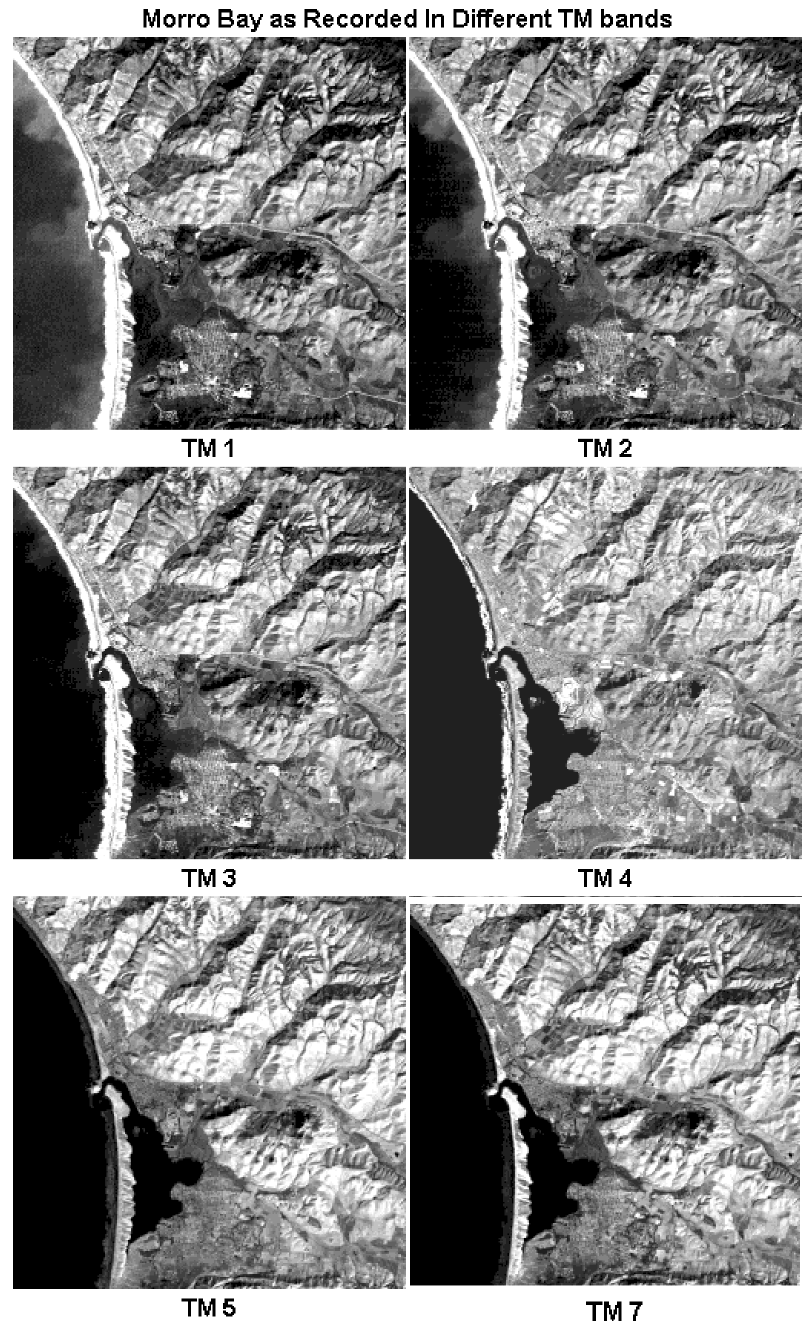
•(First three bands from the Landsat TM scene)

Let‘s have a look…

– A typical Landsat image has 7 bands
  - Blue
  - Green
  - Red
  - Near Infrared
  - Shortwave Infrared
  - Thermal Infrared
  - Shortwave Infrared2

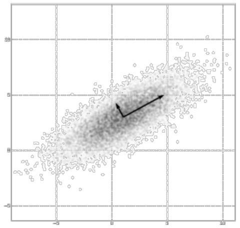
– Do these bands look really different?

– How much redundant information is there?



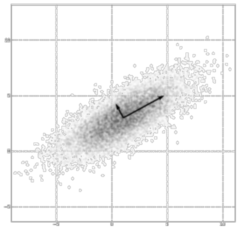
DLR

Principal Components Analysis



- *Principal Components Analysis* (PCA) is a technique used to reduce multidimensional data sets to lower dimensions
  - It describes *n*-dimensional data with a set of *p* synthetic variables, with $p < n$
  - The new variables are uncorrelated and are called *Principal Components* (PC)
  - This process leads to some information loss
    - PCA ensures that this loss is minimal
- Also known as:
  - Karhunen-Loève transform
  - Hotelling transform
  - Proper Orthogonal Decomposition (POD)

Principal Components Analysis



– PCA is widely used in remote sensing → dimensionality reduction aids data exploration

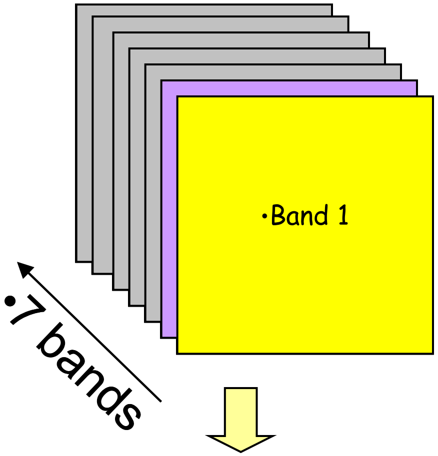
– It reveals the internal structure of the data by ignoring not relevant information

  – It highlights similarities and differences within the data

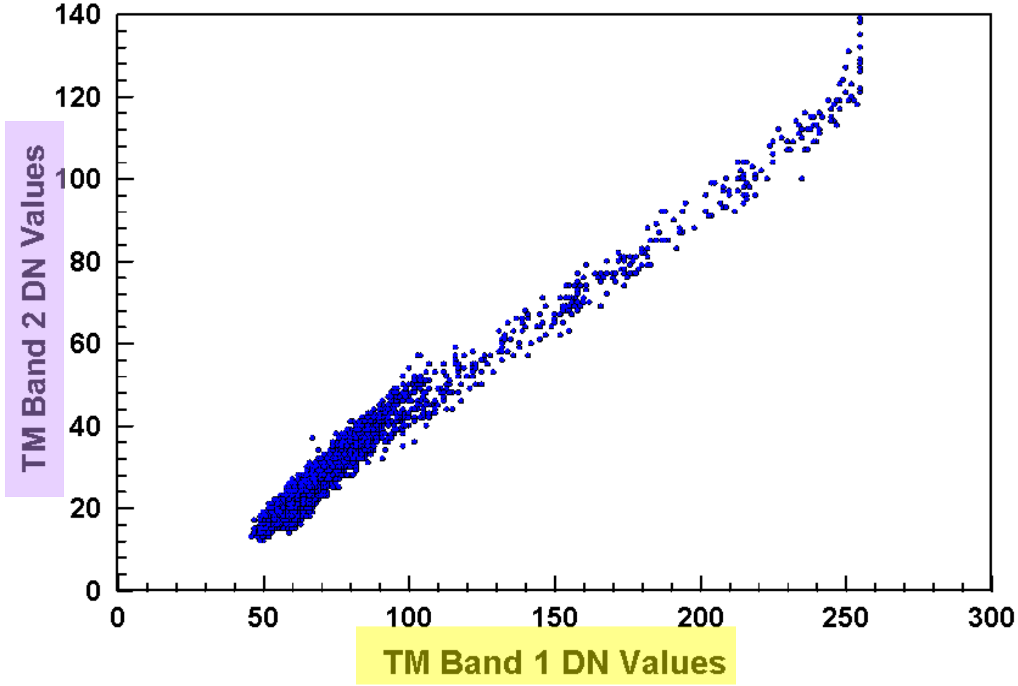
– First of all, let's see how PCA can be useful…

Let's compare some bands…

Morro Bay (California, USA) Landsat Scene



•RGB Combination



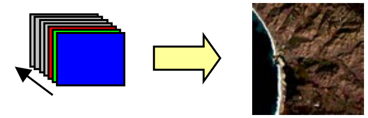
•Band 1

•Band 2

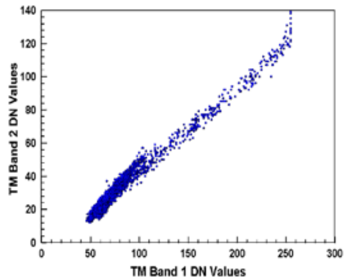
- What do you understand from the scatter plot?
- Can we predict the value of band 2 knowing band 1?

Dimensionality reduction: why?

- How to visualize multidimensional data?
  - In the previous example, out of a 7-band image only 3 bands could be visualized

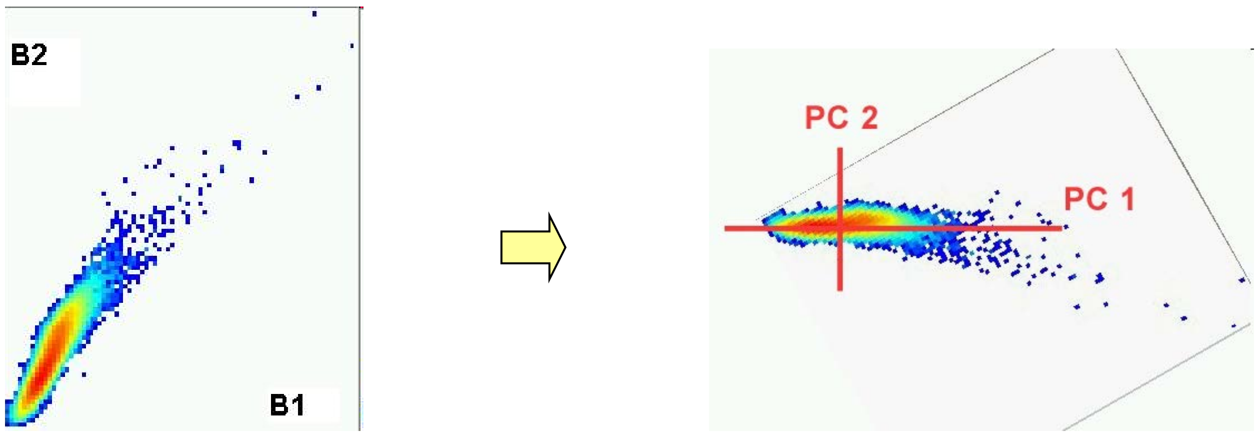


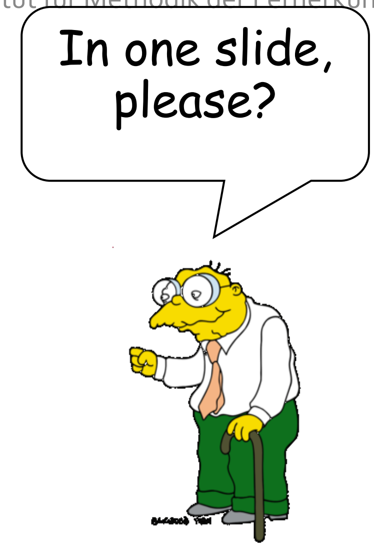
- ↑ Data → ↑ Information? Not always..
  - Redundancies
  - In the previous plot we can predict the value of band 2 on the basis of band 1



- We would like each band to contain relevant information
  - A decorrelation of the bands may help at analyzing the images

How does PCA work?





– PCA is a methodology for transforming a set of correlated variables into a new set of uncorrelated variables

– Achieved through a rotation of the original dimensions/axes to new orthogonal axes

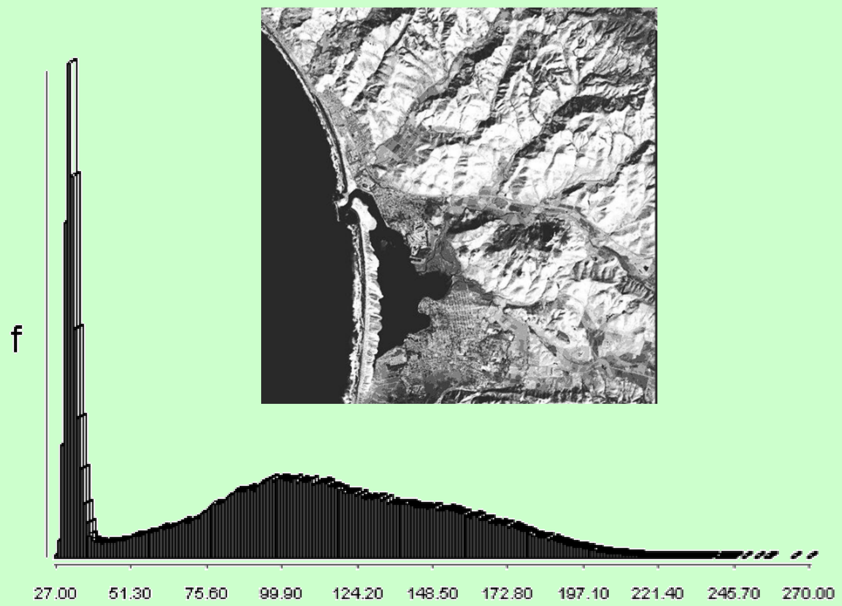
– The rotation is performed in order to have maximum variability in each new dimension

– No correlation between new variables

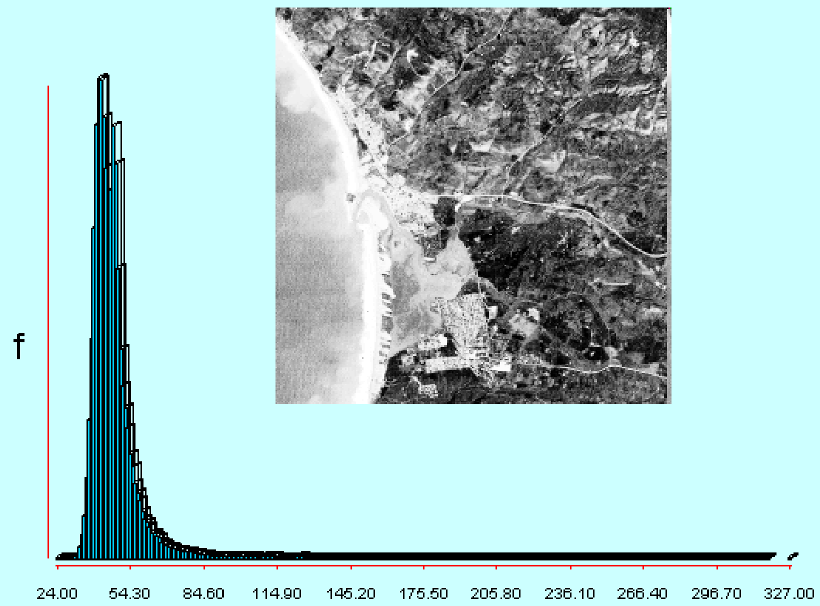
DLR

Histograms of First and Second Principal Components

Morro Bay Landsat Scene



PC1

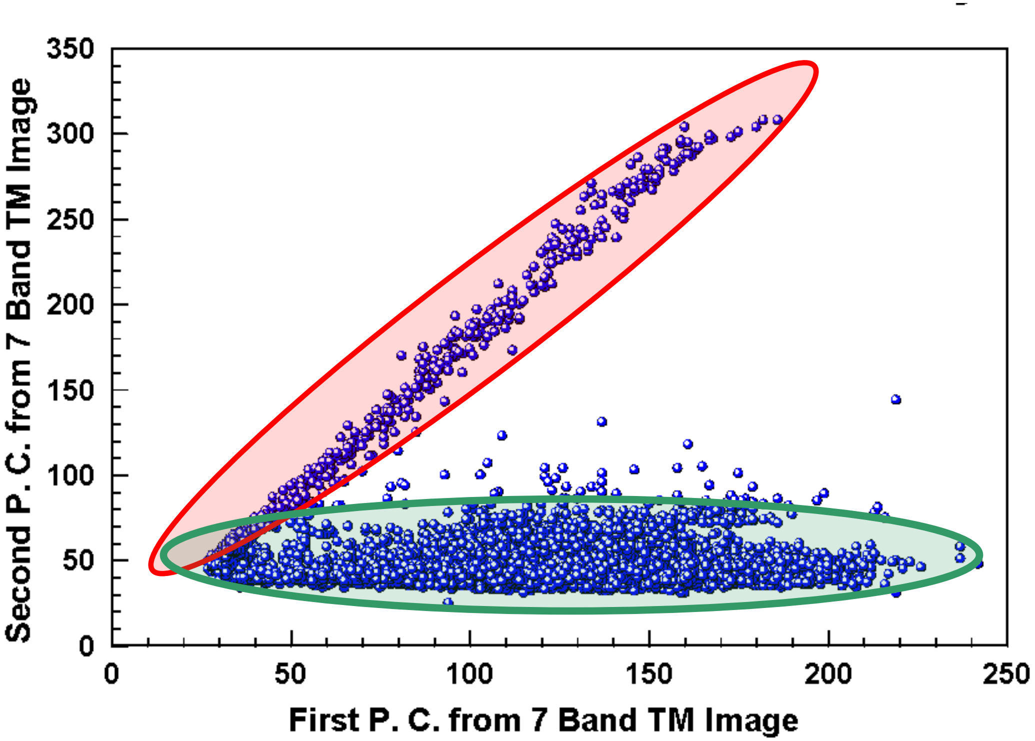


PC2

Check both the histograms and the images:

Which principal component contains more information, PC1 or PC2?

Plot of First vs. Second Principal Component



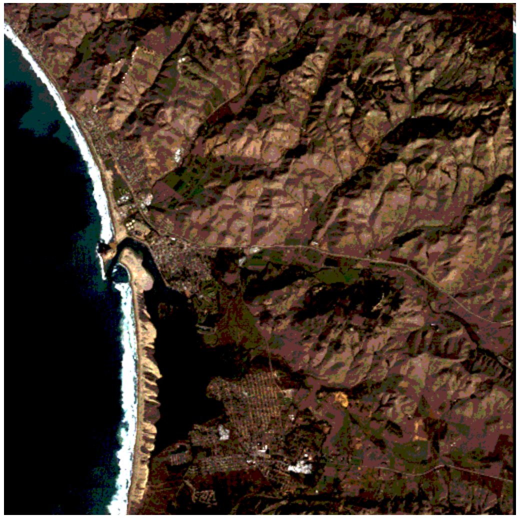
Check the scatter plot:

In which area are most of the uncorrelated data to be found?

Which pixels are still correlated?

Each component has its characteristics…

RGB Combination



PC 1

Close to what we would expect for a b/w picture of the scene

Max Information

PC 2

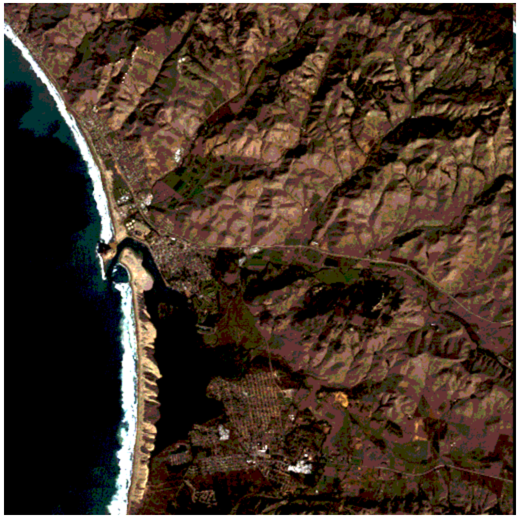
Several features can be spotted in the sea

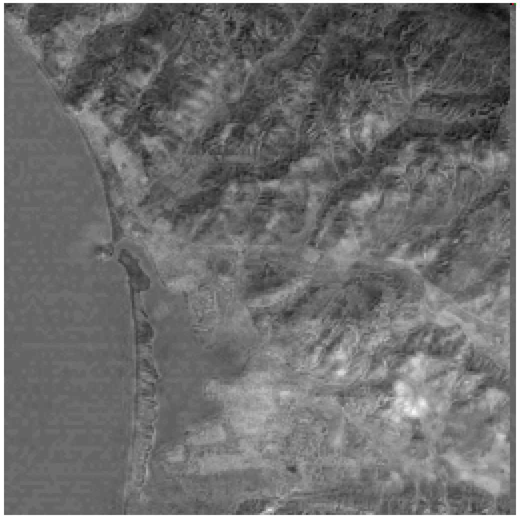
PC 3

Bright and dark gray for two classes of vegetation

Each component has its characteristics…

RGB Combination





PC 4

Still some patterns in medium gray over the mountains

PC 6

This component appears noisy

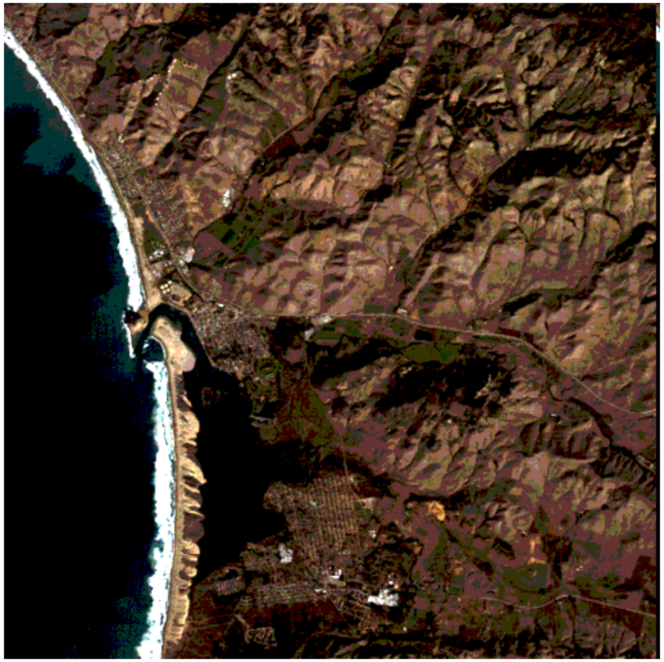
Informational content ↓

Different PC = different information!

The main keyword for PCA is…

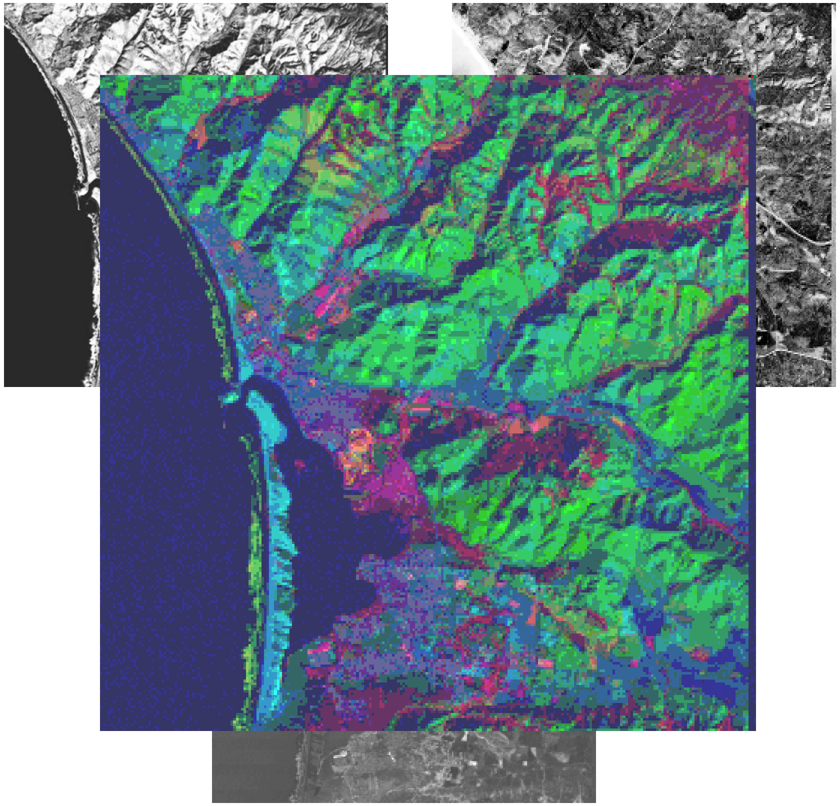
DECORRELATION!

Two Different Band Combinations



RGB Combination

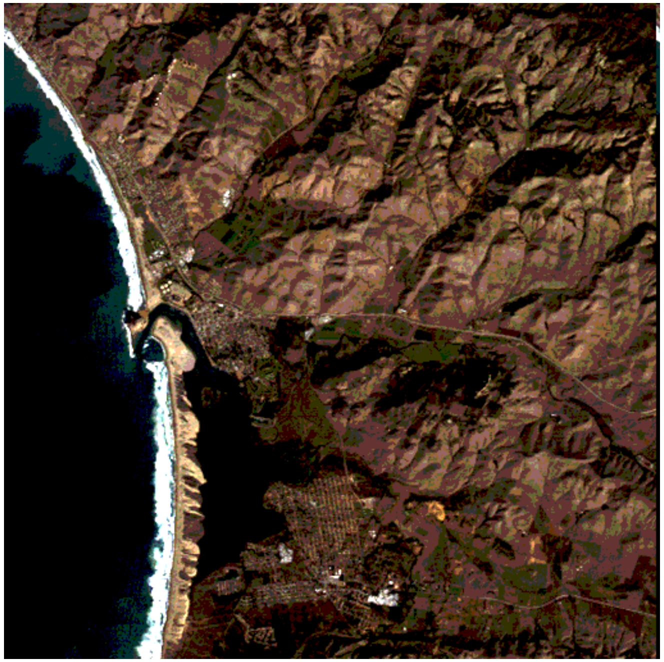
(First three bands from the Landsat scene)



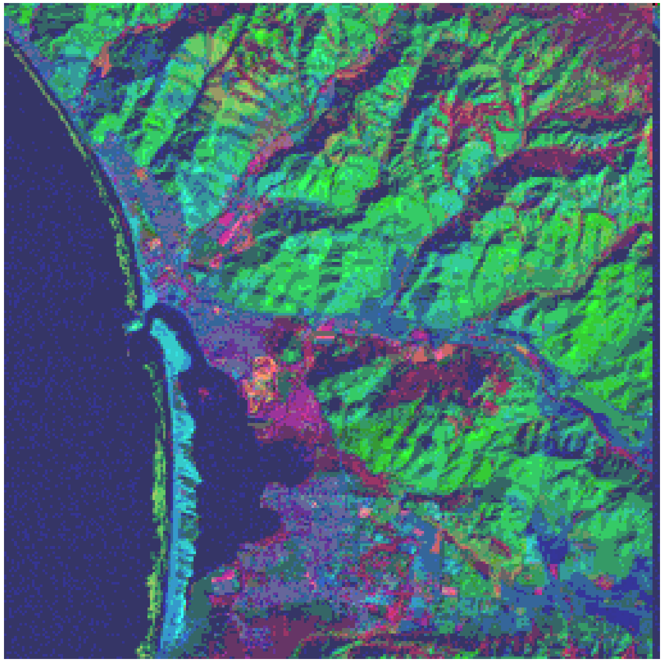
Combination of 3 PC

The information available in the Principal Components can be better revealed by combining them visually in a color composition

Two Different Band Combinations



RGB Combination

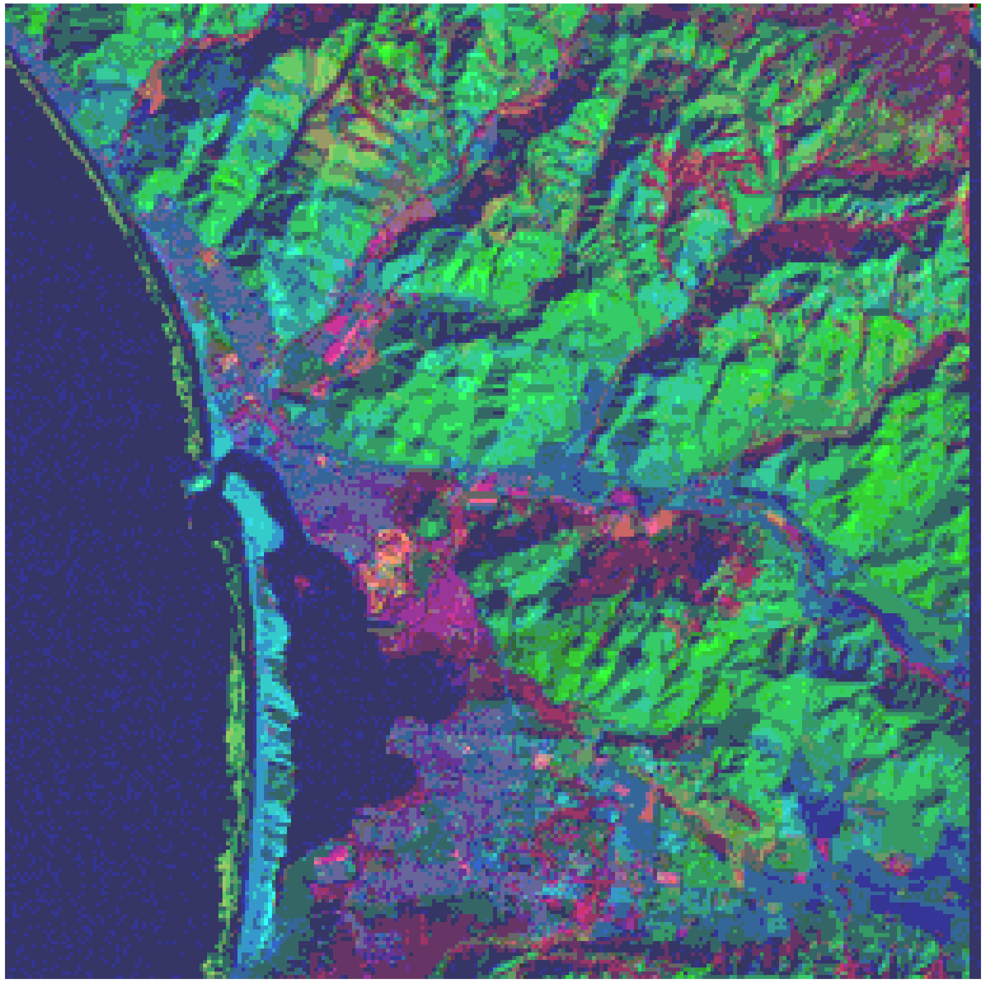


Three Principal Components

Which picture contains more information?

How many kinds of terrain can you spot in each one?

We can now identify many different areas…

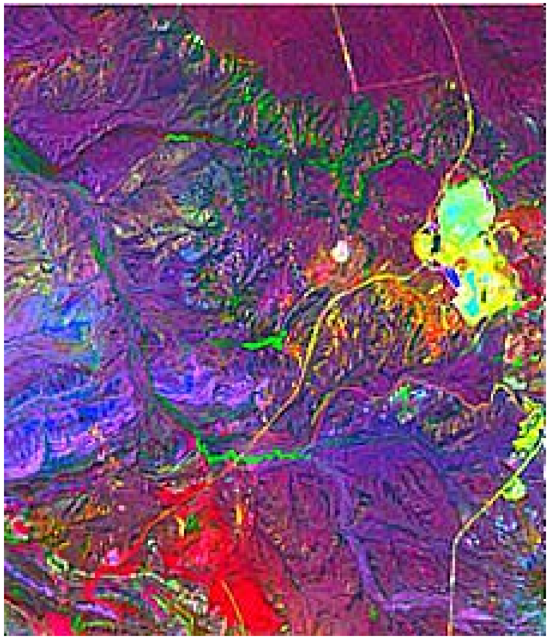


- Beach Bar
- Wave Breakers
- Vegetation1
- Vegetation2
- Golf Course
- Urban Area
- Shadows
- Sea
- Mountains (bright slopes)
- ….

A more dramatic example

A second Landsat scene

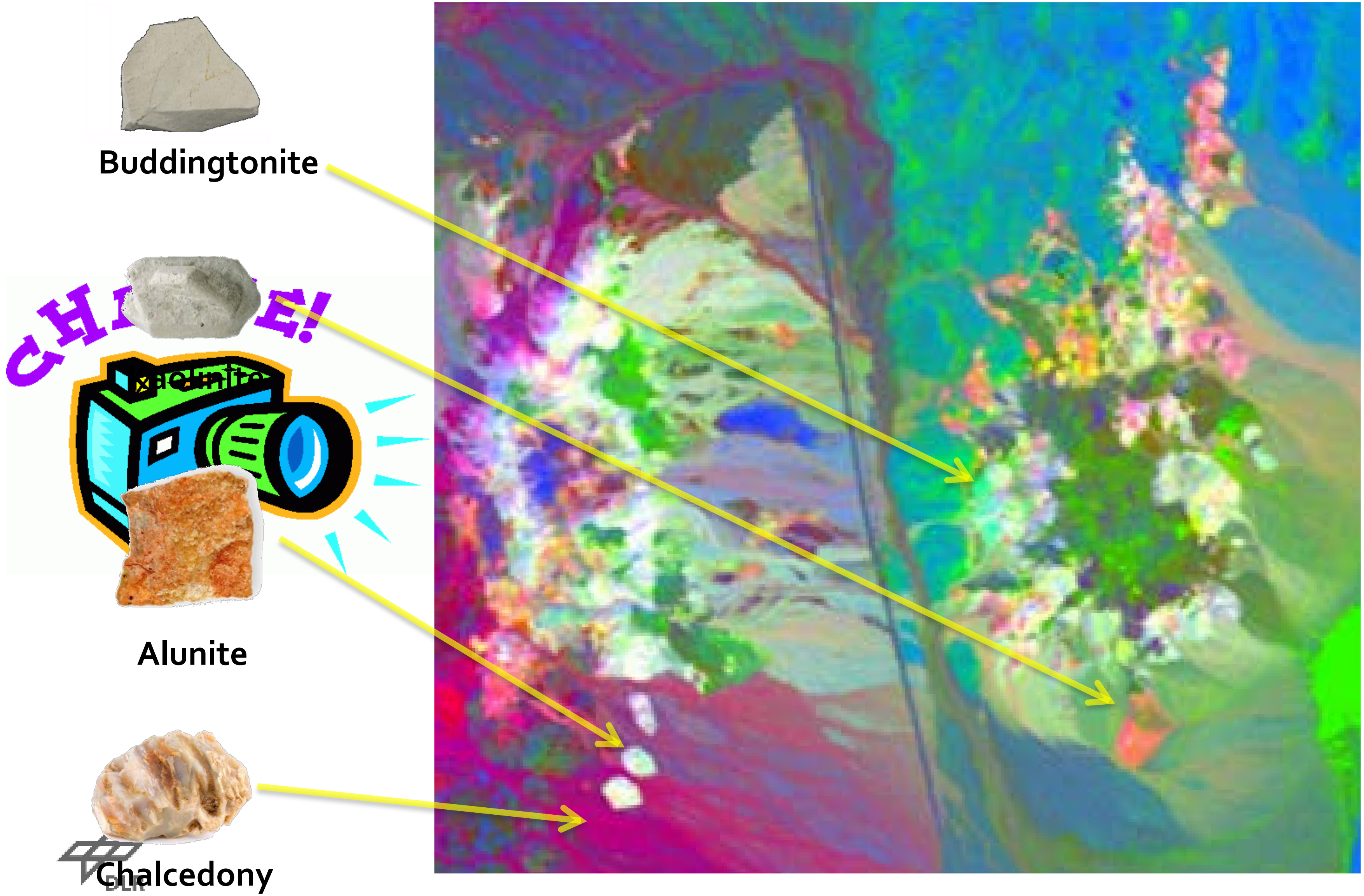
RGB Combination



Three Principal Components after Decorrelation Stretch (DS)

DS= Emphatization of the differences in color between the pixels

Hyperspectral



How do we get there?

– How do we express a principal component as a linear combination of the image bands?

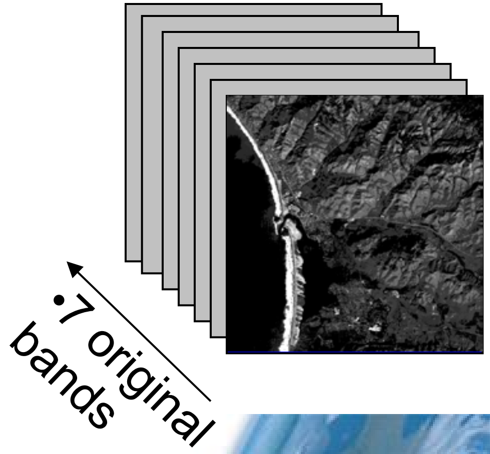
– A pixel p(i,j) at row i, column j is a vector of 7 bands b1…b7:

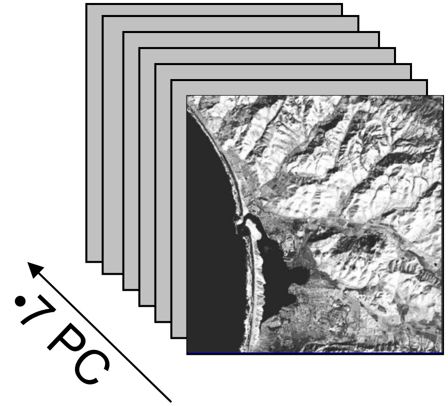
p(i,j) = [b1(i,j),b2(i,j),b3(i,j),b4(i,j),b5(i,j),b6(i,j),b7(i,j)]

– Then a pixel of a PC can be expressed as:

PC1(i,j) = [a(1,1)b1(i,j), a(1,2)b2(i,j), a(1,3)b3(i,j), a(1,4)b4(i,j), a(1,5)b5(i,j), a(1,6)b6(i,j), a(1,7)b7(i,j)]

- How can we find these a(m,n) indices for each band and each PC?

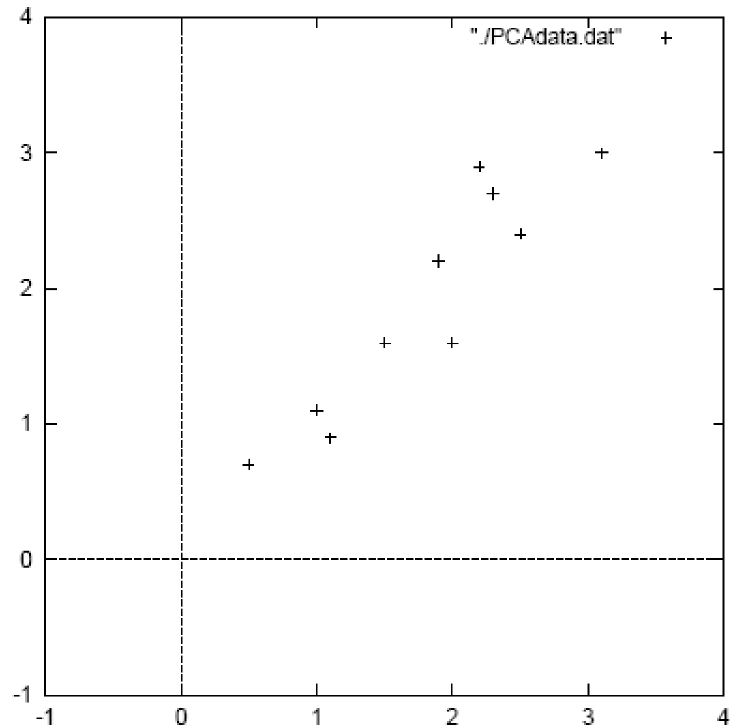




Let's see it through an Example…

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

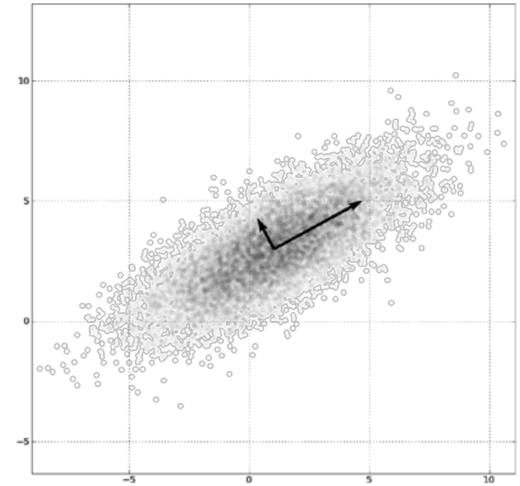


– Let's analyze this simple 2-dimensional dataset

  – Easy to visualize and to work with

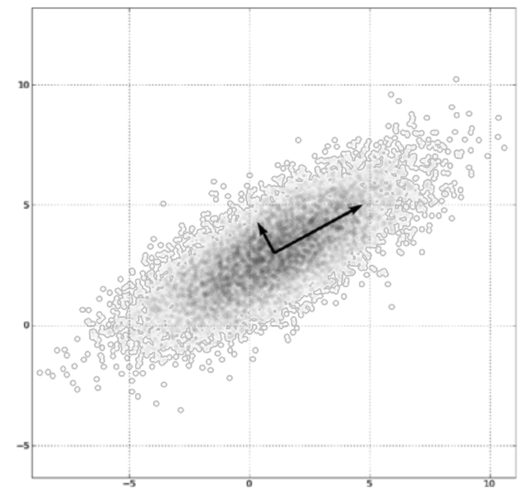
– The same procedure can be applied on the 7 dimensional Landsat scene, as well as on *n*-dimensional data (as long as *n* is finite)

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

PCA Step by Step

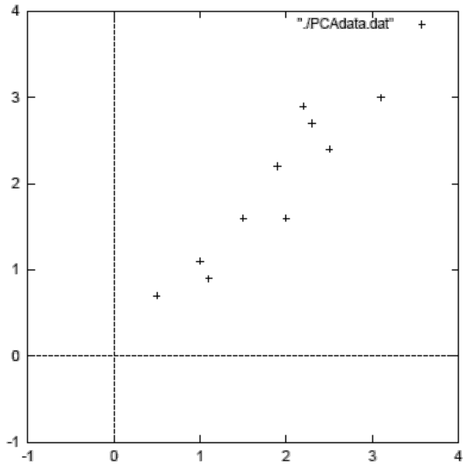


1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

Organize the Dataset

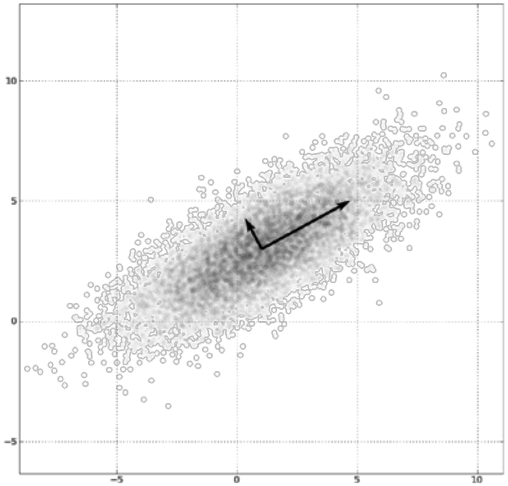
– Represent the data with a *m* x *n* matrix *M*

- m variables (in our case x and y)
- n observations per variable



$$M = \begin{pmatrix} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ 1.9 & 2.2 \\ 3.1 & 3.0 \\ 2.3 & 2.7 \\ 2 & 1.6 \\ 1 & 1.1 \\ 1.5 & 1.6 \\ 1.1 & 0.9 \end{pmatrix}$$

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

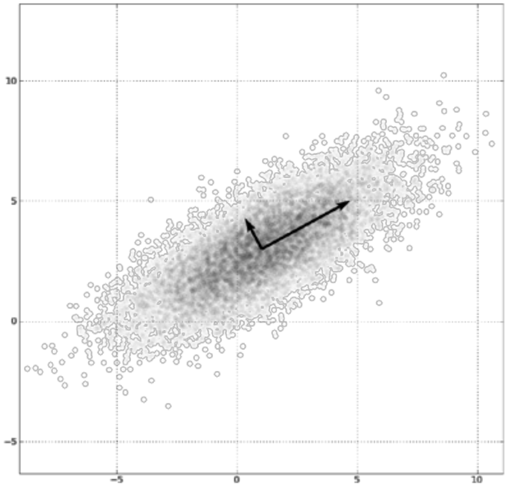
DLR

Subtract the mean

- Let $\underline{x}$ and $\underline{y}$ be the means of the *x* and *y* variables, respectively
  - For every *x* value: $x = x - \underline{x}$
  - For every *y* value: $y = y - \underline{y}$
- The mean of the data set is now zero
- Subtracting the mean makes next variance and covariance calculation easier by simplifying their equations
- The variance and co-variance values are not affected by the mean value

$$M = \begin{pmatrix} 2.5 & 2.4 \\ 0.5 & 0.7 \\ 2.2 & 2.9 \\ 1.9 & 2.2 \\ 3.1 & 3.0 \\ 2.3 & 2.7 \\ 2 & 1.6 \\ 1 & 1.1 \\ 1.5 & 1.6 \\ 1.1 & 0.9 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{pmatrix}$$

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

DLR

What is the covariance?

- The covariance *Cov*(x,y) between two variables x and y measures how much x and y change together

- There are two extreme cases:

  1. The variables are *independent*: knowing the value of x does not help in estimating the value of y → $Cov(x,y) \approx 0$

  2. The link between the variables is so strong that we can recover the values of y only by knowing the values of x → $Cov(x,y) = Max$

- Normally, this mutual dependance is somewhere in between

- High *Cov*(x,y) → High correlation → When x is positive/negative, so is y

  - If the mean of x and y has been set to 0 as in the previous example

What is a covariance matrix?

– If $\underline{x}$ and $\underline{y}$ are the mean values of x and y we can think of the covariance as the average product of the deviations of x and y from the mean:

$$Cov(x,y) = average\left[(x-\underline{x})(y-\underline{y})\right]$$

- For the 2-dimensional case we can write in a matrix the covariances of any combination of the two variables

$$CovM(x,y) = \begin{pmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{pmatrix}$$

- Where $Cov(i,i)$ is the covariance of a variable with itself
  - Better known as variance σi^2 of i

Compute the covariance matrix

$$CovM(x,y) = \begin{pmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{pmatrix}$$

– Let's focus on the non-diagonal elements

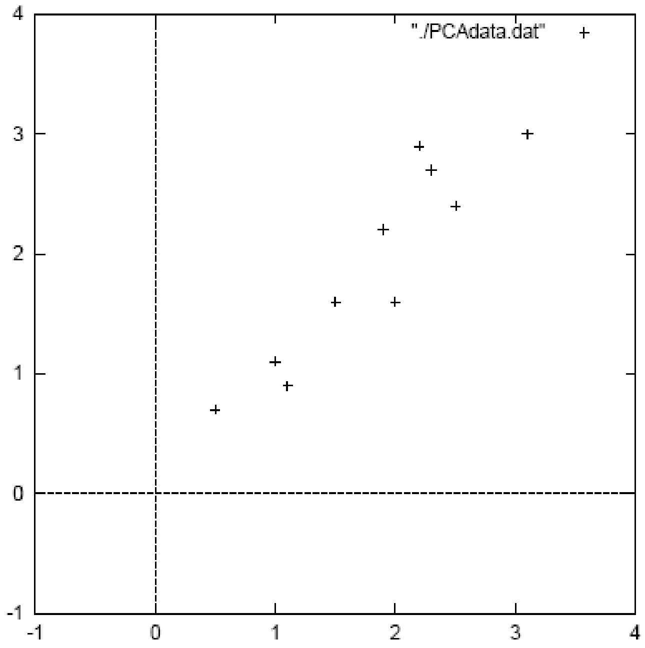
  – Related to the mutual dependence of the variables

  – This information cannot be found in the values of the diagonal containing the variances

– In this case we are interested in *Cov*(*x*,*y*)

  – It is equal to *Cov*(*y*,*x*) since the covariance matrix is always symmetric

– What kind of value do you think *Cov*(*x*,*y*) will assume for the data distribution in the figure?



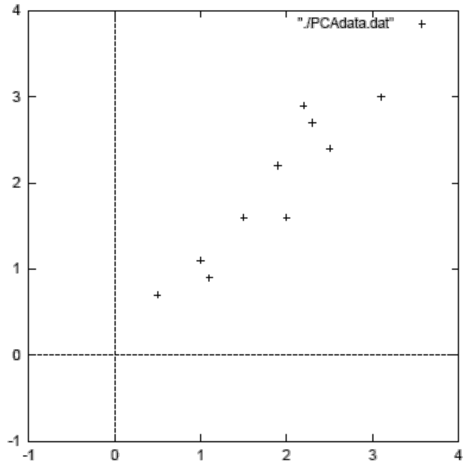
Compute the covariance matrix

$$CovM(x,y) = \begin{pmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{pmatrix} = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

– *Cov*(*x*,*y*) is positive and comparable to the variances of *x* and *y*

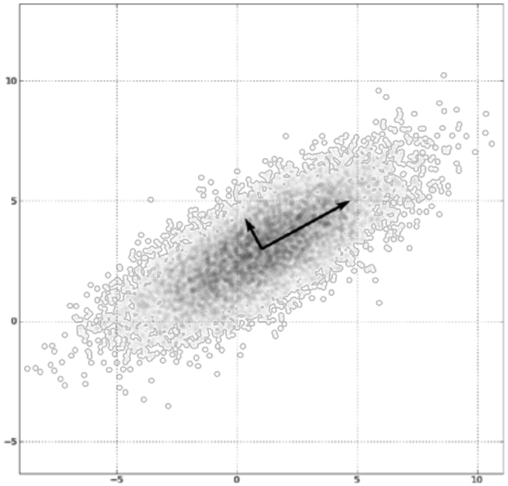
– The two variables are strongly correlated!

– We expect them to vary together



DLR

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

DLR

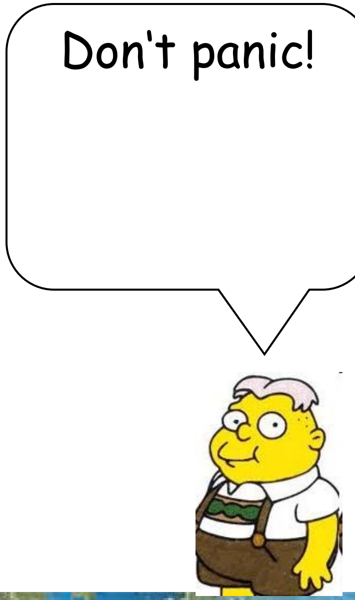
What are eigenvectors and eigenvalues?

– A vector v is an eigenvector for a matrix M if and only if

$$Mv = \lambda v$$

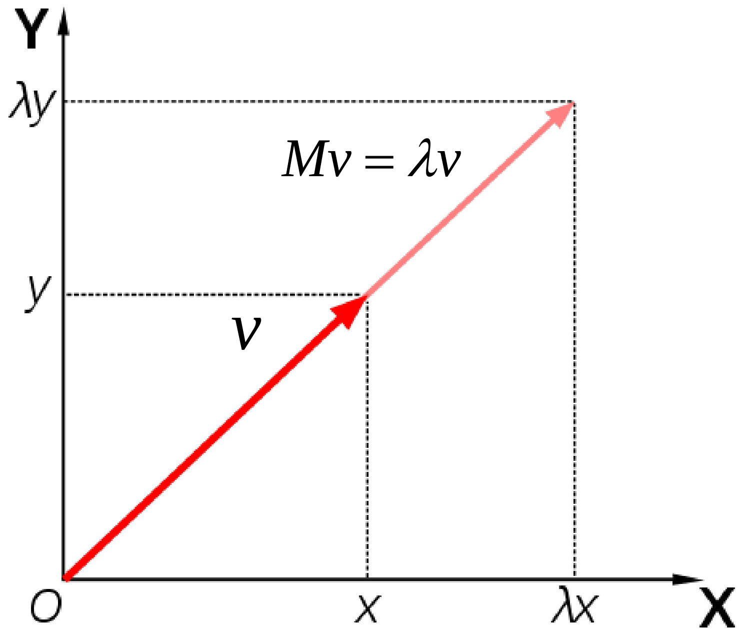
- Where λ is the eigenvalue related to the specific eigenvector v and is a scalar
- This means that v does not change if it is multiplied by M
  - The multiplication by the scalar λ „stretches“ the vector, but its direction is unaffected

- Eigenvectors are also known as characteristic vectors



DLR

What are eigenvectors and eigenvalues?



Example: here *v* is an eigenvector for the matrix *M*, as the result of the multiplication *Mv* does not change the direction of *v*.

Spot the eigenvector!

- Given the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$
- And the two vectors $x = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$ $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

x did not change after being multiplied by A!

Which one is an eigenvector?

- HINT!!

$$A\mathbf{x} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 + 1 \cdot (-3) \\ 1 \cdot 3 + 2 \cdot (-3) \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

$$Ay = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 0 + 1 \cdot 1 \\ 1 \cdot 0 + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

And what is the eigenvalue of x?

It is 1!

The vector remained unchanged

Eigenvectors and eigenvalues

– Any n x n covariance matrix A, being symmetric, has n real eigenvectors

– It can be factorized as:

$$A = Q \Lambda Q^{-1}$$

- Q → matrix composed by the eigenvectors of A
- Λ → diagonal matrix containing the eigenvalues $\lambda 1 \ldots \lambda n$
- The eigenvectors can be chosen to be orthogonal
- They can form a new orthogonal basis → they can be thought of a new set of uncorrelated variables to represent the data!

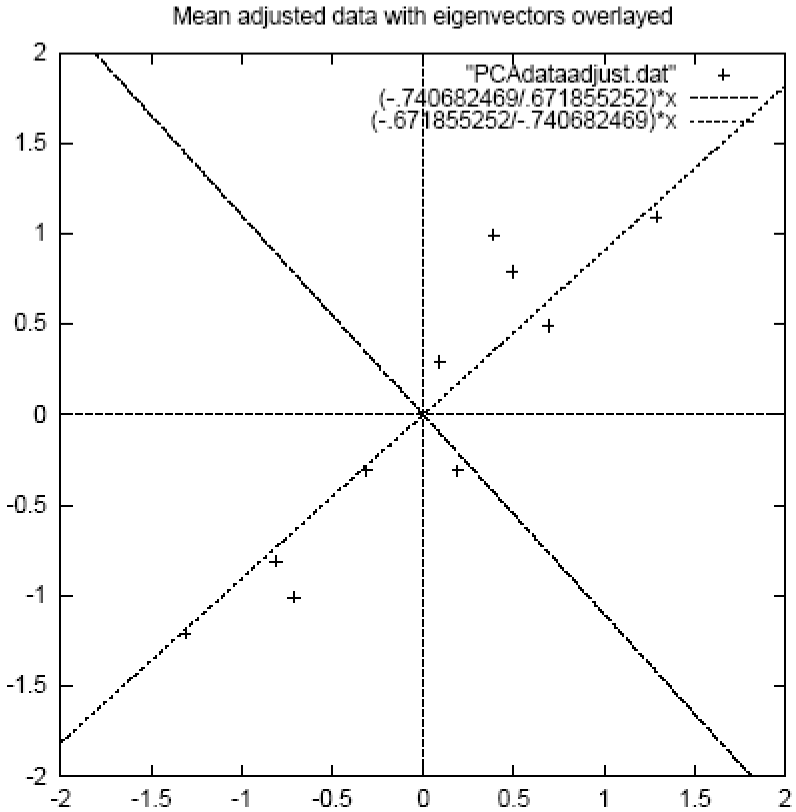
Eigenvectors and eigenvalues

– Now we can compute the eigenvectors *Q* and eigenvalues Λ for our covariance matrix…

$$CovM(x,y) = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

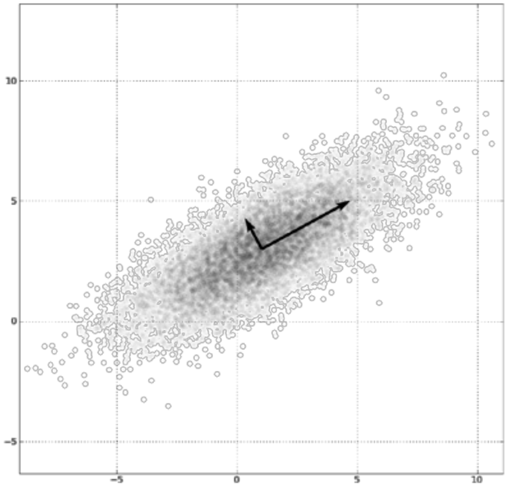
$$Q(x,y) = \begin{pmatrix} -0.735 & -0.677 \\ 0.678 & -0.735 \end{pmatrix} \qquad \Lambda(x,y) = \begin{pmatrix} 0.049 \\ 1.284 \end{pmatrix}$$

Let‘s project them back...



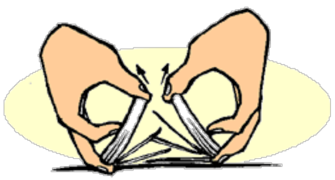
- Eigenvectors are plotted as diagonal dotted lines on the plot
- They are perpendicular to each other
- One of the eigenvectors goes through the middle of the points, like drawing a line of best fit
- The second eigenvector gives us the distance of the points from the first eigenvector
- It contains the second, less important aspect of the data

PCA Step by Step



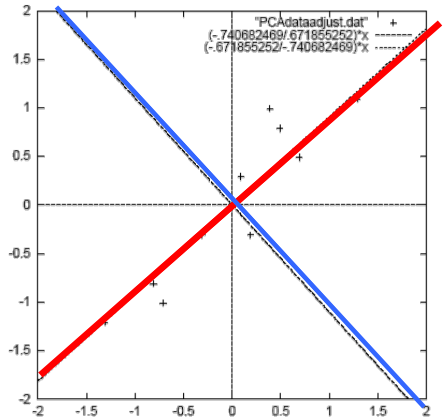
1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

Sort the eigenvectors



- The eigenvector with the highest eigenvalue is the principal component of the data set
  - It contains the highest amount of information on the data

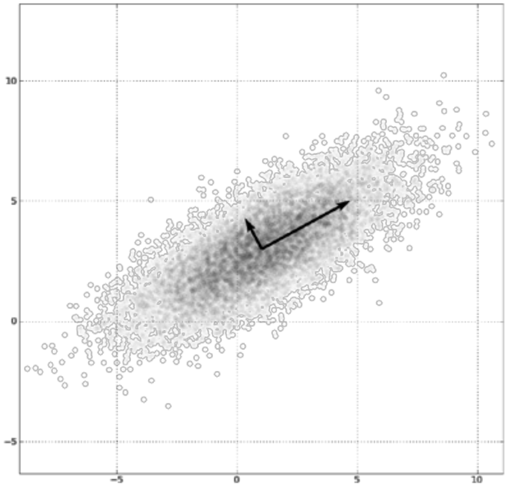
- In our example, it is "in the middle" of the data



- If we sort the eigenvectors from highest to lowest eigenvalue we have them in order of significance

$$\Lambda(x,y) = \begin{array}{cc} \begin{pmatrix} 0.049 \\ 1.284 \end{pmatrix} & \begin{array}{c} \cdot 2 \\ \cdot 1 \end{array} \end{array} \Rightarrow Q(x,y) = \begin{array}{c} \begin{array}{cc} 2 & 1 \end{array} \\ \begin{pmatrix} -0.735 & -0.677 \\ 0.678 & -0.735 \end{pmatrix} \end{array}$$

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

DLR

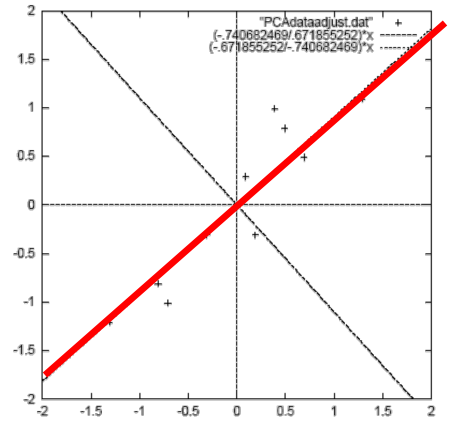
•Select a subset of the eigenvectors

– You can now decide to ignore less meaningful components

– Eigenvectors with low eigenvalue

– Dimensionality reduction is achieved

– Data compression is also achieved

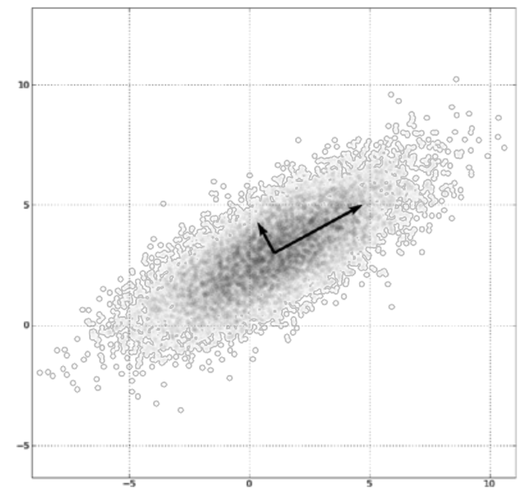


– Some information is lost, but as few as possible

$$\Lambda(x,y)=\begin{pmatrix}0.049\\1.284\end{pmatrix}\begin{matrix}\cdot 2\\\cdot 1\end{matrix}\quad\Longrightarrow\quad Q(x,y)=\begin{array}{c}\begin{matrix}\;\;2 & \qquad 1\end{matrix}\\\begin{pmatrix}-0.735 & \boxed{\begin{matrix}-0.677\\-0.735\end{matrix}}\\0.678 & \end{pmatrix}\end{array}$$

•We can choose only the first component!

PCA Step by Step



1. Organize the data set
2. Subtract the mean
3. Compute covariance matrix
4. Find eigenvectors and eigenvalues for the covariance matrix
5. Sort the eigenvectors
6. Select a subset of the eigenvectors as basis vectors
7. Project the values unto the new basis

•Deriving the new data

– We can multiply our old data by our chosen set of eigenvectors

– We obtain a new representation for the 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

•New representation of the data using both PCs

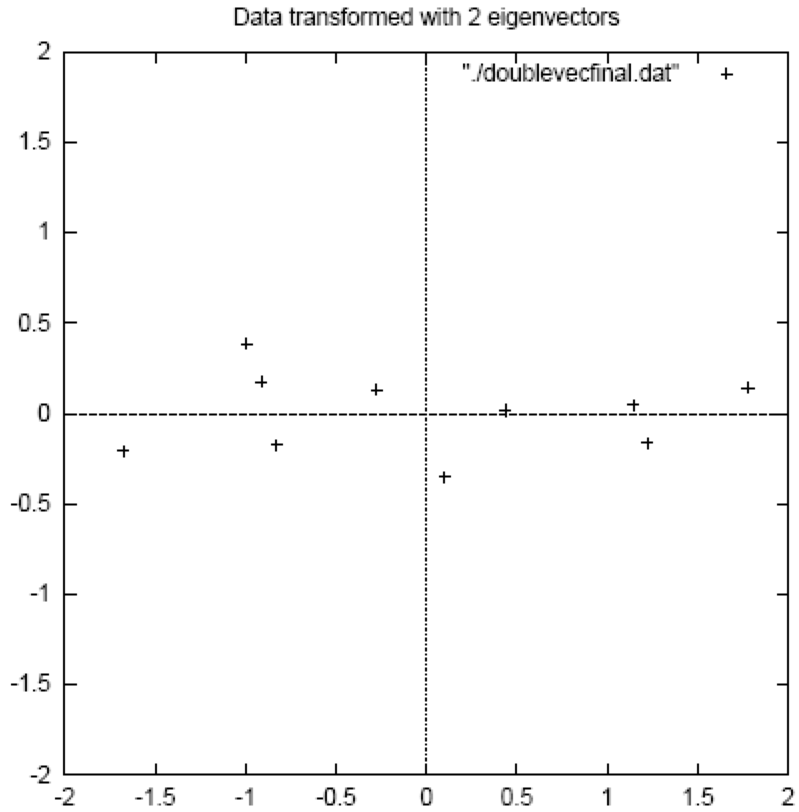


Figure 3.3: The table of data by applying the PCA analysis using both eigenvectors, and a plot of the new data points.

What if we use only the first PC?

x

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

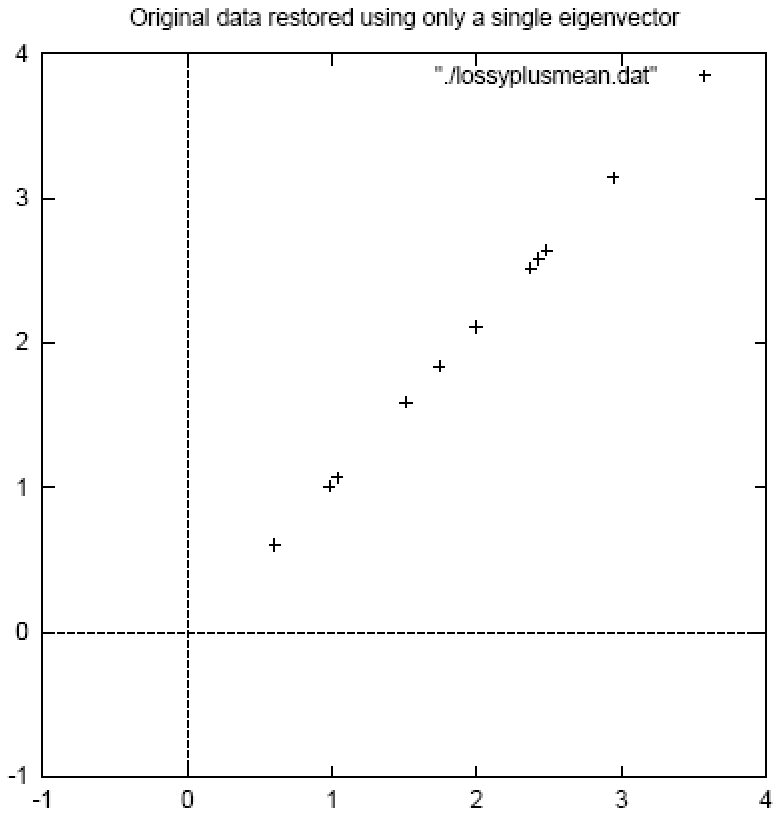


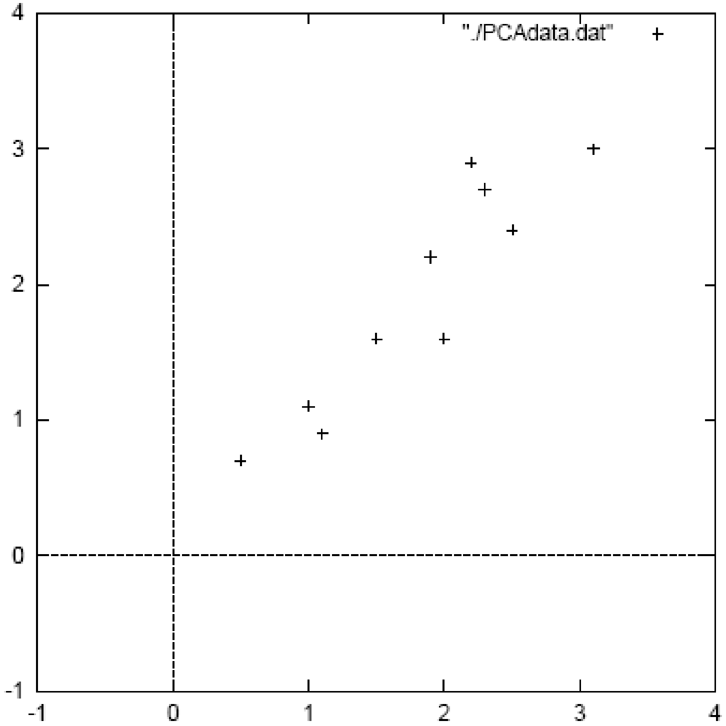
Figure 3.5: The reconstruction from the data that was derived using only a single eigen-vector

After adding back the mean values subtracted in the first steps

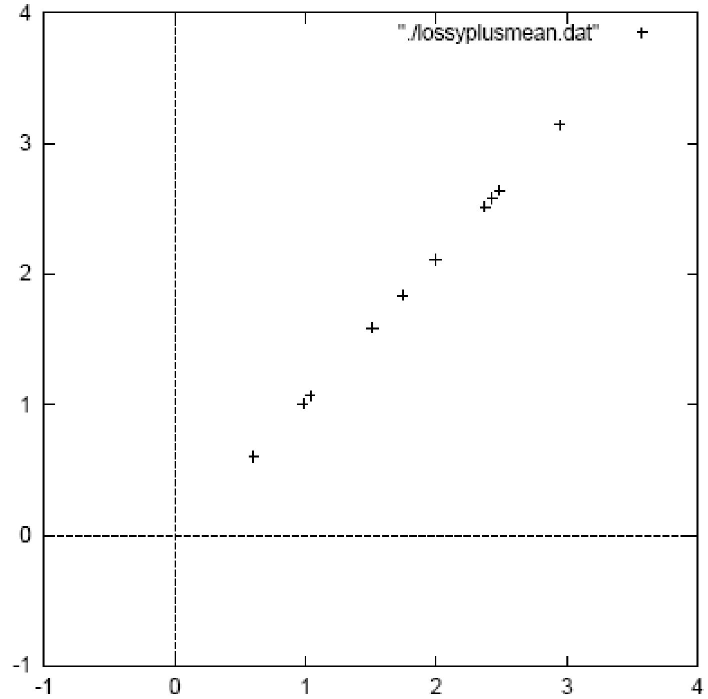
←

DLR

How much information are we keeping?

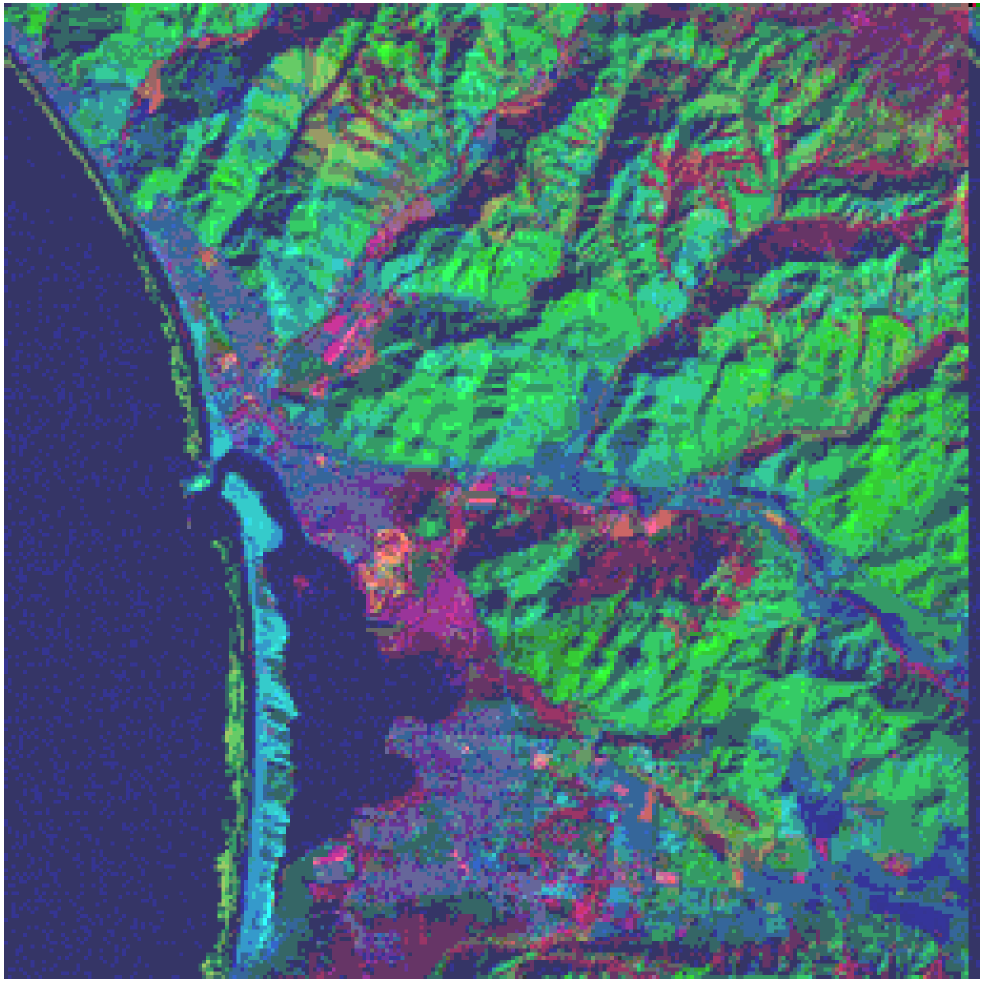


•Original 2D Data



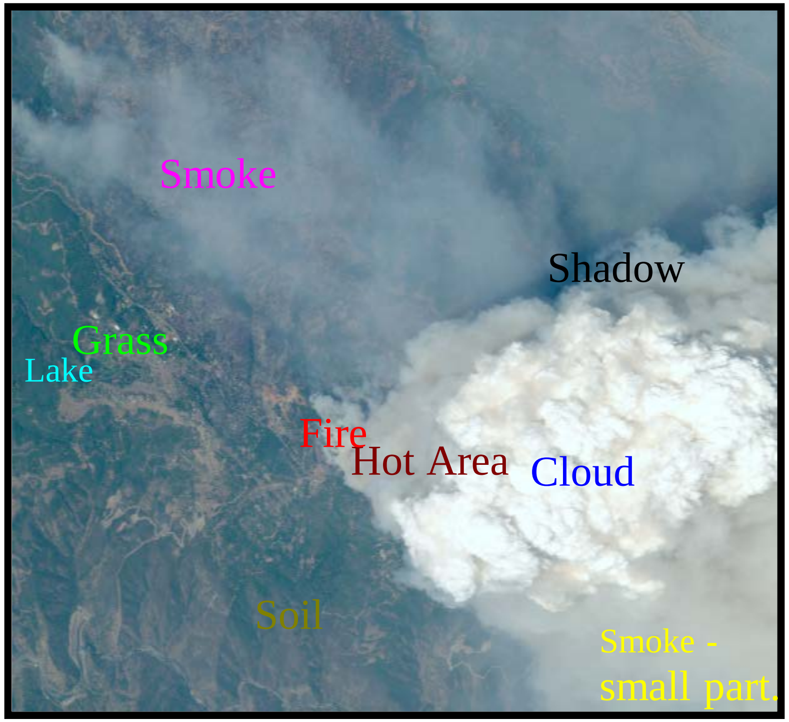
•Data reconstructed on the basis of only 1 Principal Component

•That‘s how we got here!



- •Beach Bar
- •Wave Breakers
- •Vegetation1
- •Vegetation2
- •Golf Course
- •Urban Area
- •Shadows
- •Sea
- •Mountains (bright slopes)
- •....

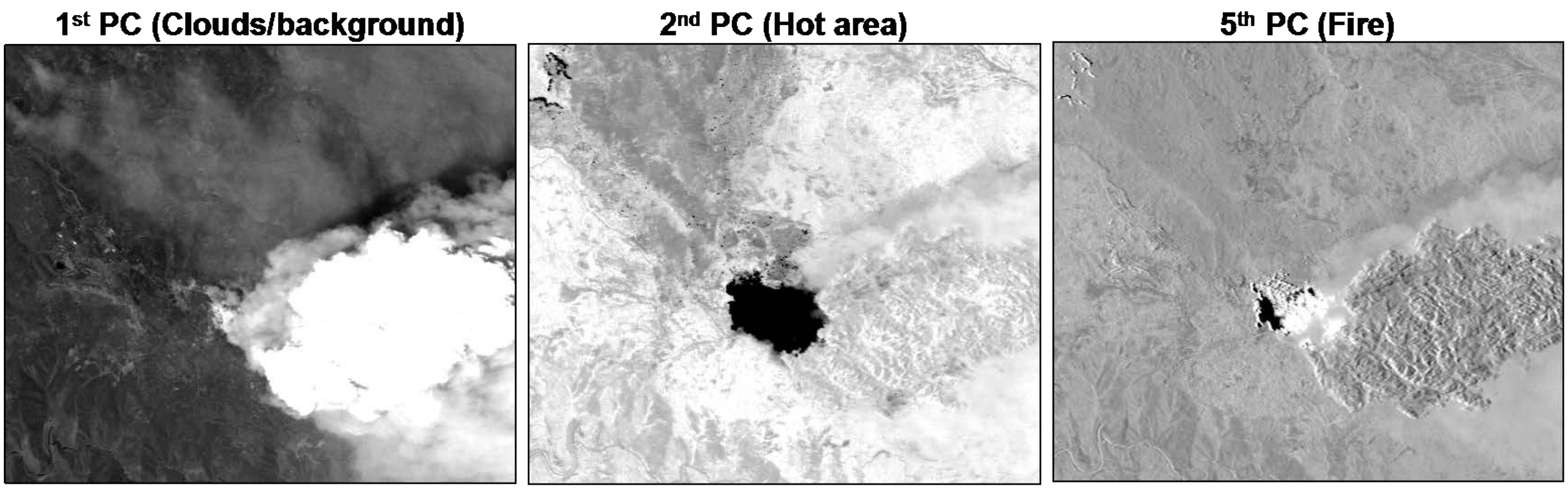
One Last Example



AVIRIS sensor RGB, Linden, CA , 20-Aug-1992

(Hsu, et al. in Frontiers of Remote Sensing Information Processing, WSP 2003)

Three Principal Components

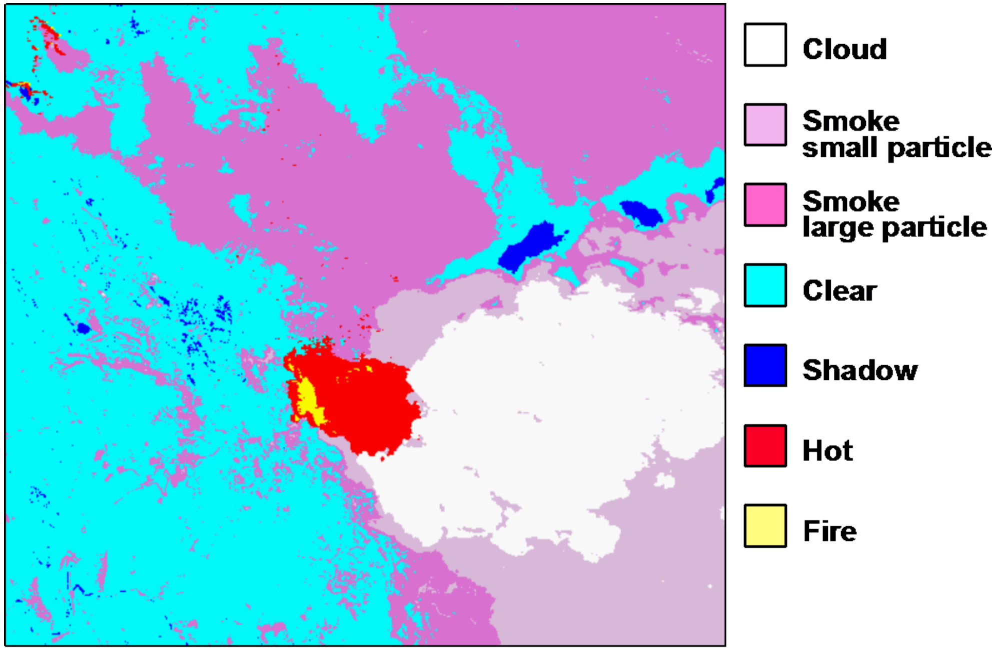


The 1st component again resembles a b/w picture of the area

The 2nd highlights an area in which we have a thermal anomaly

The 5th shows the cause of the anomaly (fire), which was hidden in the true color composition

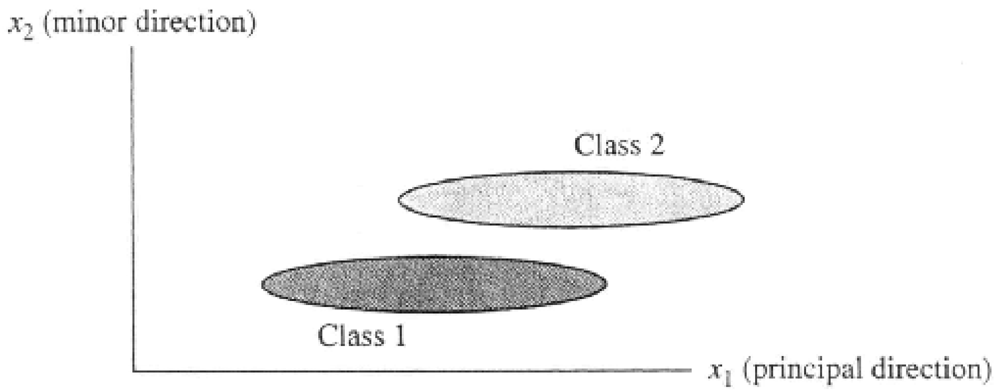
Classification using the 3 PCs



All major atmospheric and surface features can be identified

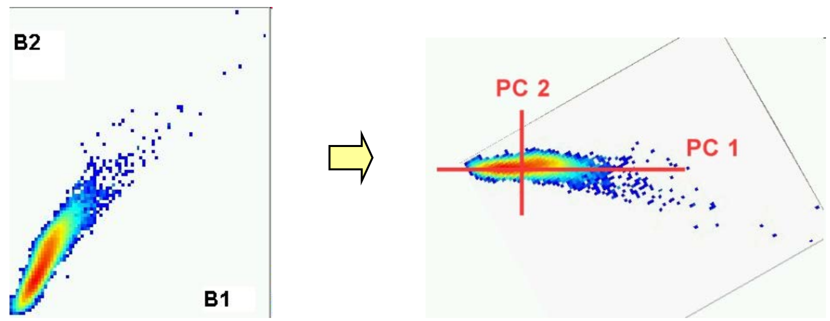
DLR

PCA is NOT Always Optimal!



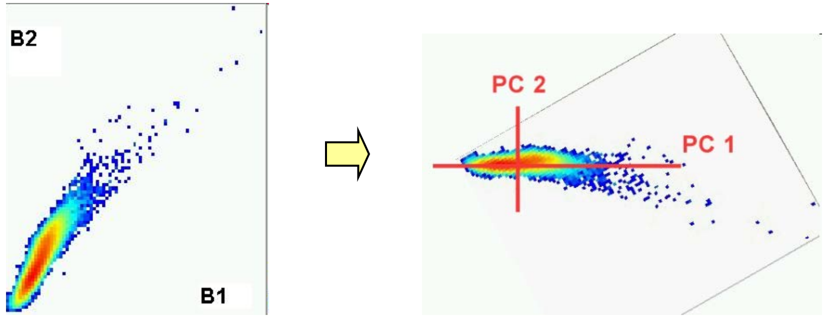
– What happens if *x1* and *x2* are our first two PCs in this example?

Questions



- What is the relation between the eigenvectors of the covariance matrix and the principal components?
- At what point in the PCA process can we decide to compress the data?
- Why are the principal components orthogonal?
- How many different covariance values can you calculate for an n-dimensional data set?

Conclusions



– PCA can be viewed as a projection of the observations onto orthogonal axes contained in the space defined by the original variables

– The first new variable (PC1) contains the maximum amount of variation → max information

– The remaining components PC2..PCn are sorted according to their informational content, i.e. to their variance (which is not equal to the variance of the variables!!)

– The rotation is a linear combination of the original bands

→ No information loss, original data can be recovered

→ The last components can be ignored, achieving data reduction

DLR