Principle component analysis

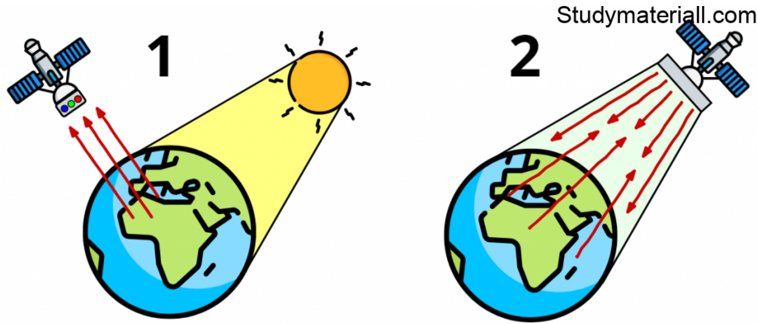
Karina Nielsen (karni@space.dtu.dk) and Allan Aasbjerg Nielsen

September, 2022

Overview

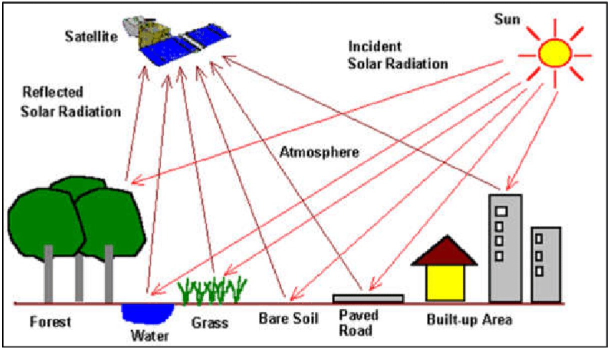
- Repetition from last week
- Multivariate normal distribution and linear transform
- What is principal component transform?
- Derivation of PCA
- What is it used for?
- Examples
- Exercises (Report, not mandatory, but strongly recommended to do)
- Change: 13/9 Daniel will be lecturing on his first topic

Principle of Remote Sensing



Optical Remote Sensing

- Capture the reflected sun light
- The various materials will reflect the sun light differently



Reflectance and wavelength

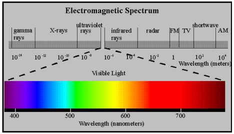


Figure 1: Source: Paul R. Baumann

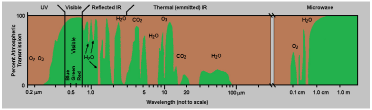


Figure 2: Source: Paul R. Baumann

- Some wavelength are absorbed in the atmosphere
- The reflectance of a material change as a function of wavelength
- The difference in reflectance of materials as a function of wavelength makes it possible to distinguish between surface materials

Examples

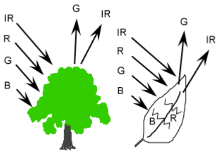


Figure 3: Health vegetation

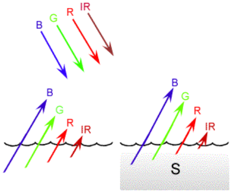


Figure 4: Water

Examples

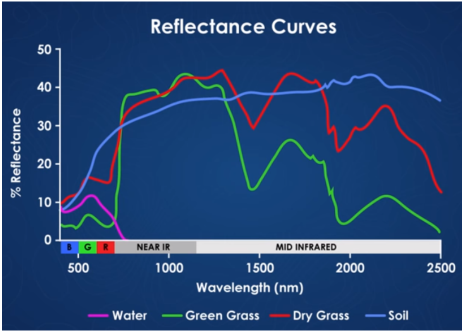


Figure 5: Source: National Ecological Observatory Network (NEON)

Exercises from last week

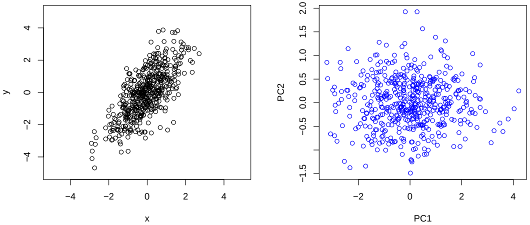
- Spectral response curves
- Normalized burn ration
  - Consult the following web-page to learn and understand the NBR https://un-spider.org/advisory-support/recommended-practices/recommended-practice-burn-severity/in-detail/normalized-burn-ratio
- Links to solutions
  - Spectral response curves: https://code.earthengine.google.com/60b06807cfe758339036930b016d1b4f
  - https://code.earthengine.google.com/?scriptPath=users%2Ficefollower%2Fmyfirstscript%3Aexample3burn

Intuitive description of PCA

- A linear transform, that transform the correlated variables Z into uncorrelated factors Y

$$Y = ZA$$

- Z is an $n \times m$ matrix where m is the number of features (e.g. bands) and n is the number of observations.
- A is an $m \times m$ matrix



- This means that the covariance matrix of Y must be a diagonal matrix
- Now we just need to find the rotation matrix A

When we have correlation between data, we have less info than we think Because we can predict one with the other

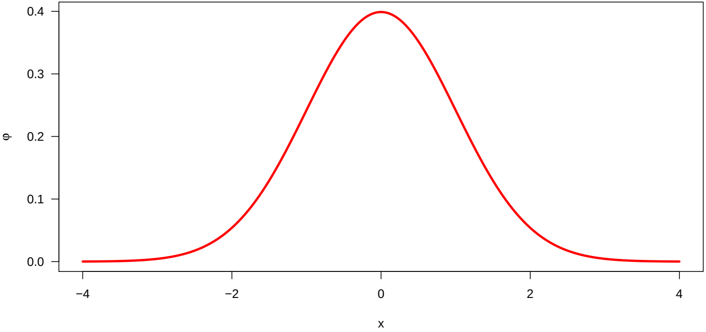
- Understand difference between uncorrelated and independent
- PCA to normal distribution gives out independent and uncorrelated variables
- PCA is very much to reduce redundancy in the dataset

The Normal Distribution - a few facts

- A continuous probability distribution on $(-\infty, \infty)$
- The probability density function is:

$$\varphi(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- The mean is μ and standard deviation is σ
- The interval $(\mu - 2\sigma, \mu + 2\sigma)$ contains 95%

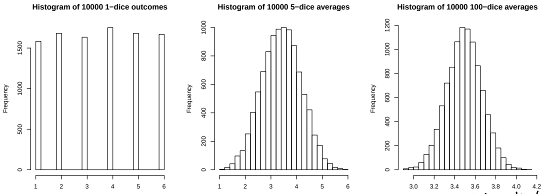


Central Limit Theorem (CLT)

- If $X_1, X_2, \ldots$ are independent identically distributed variables with finite mean μ and variance σ^2, then

$$\sqrt{n}\frac{\frac{1}{n}\sum X - \mu}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ , as } n \to \infty$$

- Notice that nothing is said about the distribution of X (except about mean and variance)



no matter which distribution

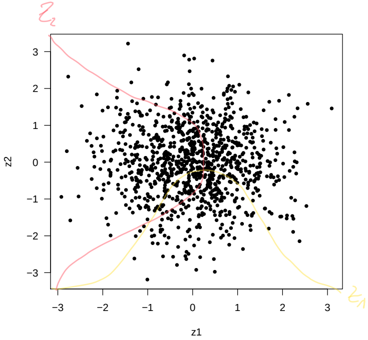
- Why is this important? If we have enough data, the average will be normal distributed! To which we can apply PCA with nice results

Two normal random variables

- Imagine we have two univariate normally distributed random variables

$$Z_1 \sim N(0,1) \text{ and } Z_2 \sim N(0,1)$$

- If we plot a lot of simulations (z_1, z_2) we get:



- The marginal distribution on each axis is a $N(0,1)$

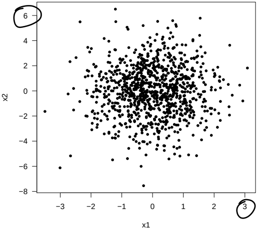
Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$X = \begin{pmatrix} Z_1 \\ 2Z_2 \end{pmatrix}$$

more spread out

- If we plot a lot of simulations (x_1, x_2) we get:



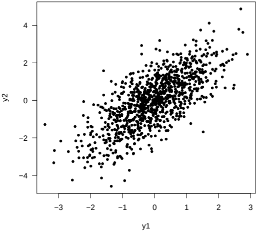
- The marginal distribution is a $N(0, 1)$ on first axis and $N(0, 4)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$Y = \begin{pmatrix} Z_1 \\ Z_1 + Z_2 \end{pmatrix}$$

- If we plot a lot of simulations (y_1, y_2) we get:



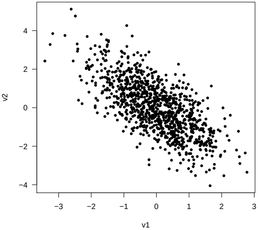
- The marginal distribution is a $N(0, 1)$ on first axis and $N(0, 2)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$V = \begin{pmatrix} Z_1 \\ Z_2 - Z_1 \end{pmatrix}$$

- If we plot a lot of simulations (v_1, v_2) we get:



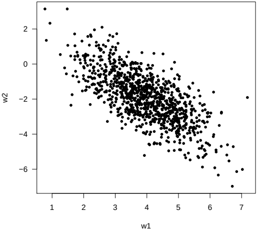
- The marginal distribution is a $N(0,1)$ on first axis and $N(0,2)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$W = \begin{pmatrix} Z_1 + 4 \\ Z_2 - Z_1 - 2 \end{pmatrix}$$

- If we plot a lot of simulations (w_1, w_2) we get:



- The marginal distribution is a $N(4, 1)$ on first axis and $N(-2, 2)$ in the second axis.

Two normal random variables

- If we define Z as:

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

- Then we can write all the cases as:

$$AZ + b$$

- where A is a matrix and b is a vector.
- E.g. the last example:

$$W = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} Z + \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

Multivariate normal distribution

- We say that the k-dim random variable X follows a multivariate normal distribution $X \sim N_k(\mu, \Sigma)$ if there exsists random l-dim random varible Z where each component follows a $N(0,1)$ distribution, such that $X = AZ + b$.
- In that case $\Sigma = AA^t$ and $\mu = b$

Which is often the case in spatial data

- The density for a k-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ is:

$$L(x) = \frac{1}{(2\pi)^{k/2}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

- We write $X \sim N_k(\mu, \Sigma)$.

Covariance and correlation

- The covariance between two random variables is defined as:

$$\text{cov}(X, Y) = E\left((X - \overset{\text{mean}}{\mu_x})(Y - \mu_y)\right)$$

- For a multivariate normal $X \sim N_k(\mu, \Sigma)$ we have arranged all the covariances in the matrix Σ, such that:

$$\Sigma_{ij} = \text{cov}(X_i, X_j)$$

- The covariance between a variable and itself is the variance of that variable, so

$$\Sigma_{ii} = \text{cov}(X_i, X_i) = \text{var}(X_i)$$

Valori sulla diagonale con solo la varianza

- The correlation coefficient is defined as:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

- Mini exercise: Find the correlation coefficient of W on previous page.
- Mini exercise: Can you construct an A, such that $\rho = 0.9$? Cholesky decomposition

$$\Sigma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

$$\rho_{\substack{11 \\ ij}} = \frac{1}{\sqrt{1 \cdot 1}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

$$\rho_{12} = \frac{-1}{\sqrt{1 \cdot 2}} = -\frac{1}{\sqrt{2}} = -\frac{\sqrt{2}}{2}$$

$$\rho_{21} = \frac{-1}{\sqrt{2 \cdot 1}} = -\frac{\sqrt{2}}{2}$$

$$\rho_{22} = \frac{2}{\sqrt{4}} = 1$$

$$\rho = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & 1 \end{pmatrix}$$

Linear transformation of multivariate normal

- Assume:

$$X \sim N(\mu, \Sigma)$$

- The distribution of

$$(AX + b) \sim N(A\mu + b, A\Sigma A^t)$$

Matrix deviations of PCA

- In PCA we require the transformation to be linear and orthogonal

$$Y = ZA \quad \text{with} A^T A = I$$

- The matrix A is an $m \times m$ and orthogonal
- Assume that Z represents the data minus its mean, then V is the covariance matrix of Z represented by the sample covariance S

$$V = [s_{ij}] = \frac{1}{n-1} Z^T Z$$

- The PCA states that the transformed data must be uncorrelated, which mean that the off-diagonal elements in the covariance matrix D are zero. If the data further is normal distributed the transformed data will also be independent.

$$D = \frac{1}{n-1} Y^T Y = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_{mm} \end{bmatrix}$$

Matrix deviations of PCA

- Now multiply $Y = ZA$ with $\frac{1}{n-1} Y^T$ on both sides

$$\frac{1}{n-1} Y^T Y = \frac{1}{n-1} Y^T Z A$$

- Replace Y with ZA

$$\underbrace{\frac{1}{n-1} (ZA)^T (ZA)}_{D} = \frac{1}{n-1} A^T Z^T Z A = A^T \underbrace{\frac{1}{n-1} (Z^T Z)}_{?} A$$

PCA continued

$$D = A^T VA$$

$$VA = AD$$

- This can be recognized as the eigenvalue equation, where D is the eigenvalues, A is the matrix containing the eigenvectors and V is the data covariance matrix
- An eigenvector is a vector that does not change direction during the linear transformation
- The eigenvalues and vectors can be found via eigenvalue decomposition or singular value decomposition (SVD) of V
- In PCA the eigenvectors gives the direction where the largest variance in the data is observed and the eigenvalues gives the size of the variance

Summary of the PCA transform

- Standardize the d-dimensional dataset
- Construct the data covariance matrix
- Decompose the covariance matrix into its eigenvectors and eigenvalues.
- Sort the eigenvalues by decreasing order to rank the corresponding eigenvectors.
- Construct a projection matrix A from all the (or the "top" k) eigenvectors.
- Transform the m-dimensional input dataset X using the projection matrix A to obtain the new k-dimensional feature subspace, where k is equal to or smaller that m

Interpretation of the eigenvalues and eigenvectors

- The eigenvalues gives the variance in the direction of the eigenvectors
- We can calculate the scores (how much of the total variance they each describe) of the respective eigenvalues

$$score = \frac{\lambda_i \times 100}{\sum_{i=1}^{m} \lambda_i} \quad \text{where } m \text{is the number of eigenvalues}$$

- The relation between the original bands and the PCs are described via the loads

$$R_{km} = \frac{a_{km} \times \sqrt{\lambda_m}}{\sqrt{Var_k}}$$

- Here a_{km} is eigenvector component related to the band k and PC component m
- λ_m is the mth eigenvalue
- Var_k is the variance of the kth band in the data covariance matrix. If standardized this is 1

Comments regarding the PCA

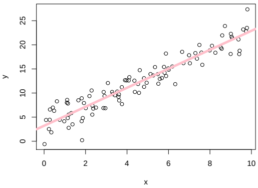
- The direction/sign of the eigenvectors is arbitrary
- The PCA is scale dependent (standardization of the data or not)
- The PCs does not have a direct physical interpretation but are linear combinations of the original bands

Common applications of PCA

- Widely used method to reduce the dimension of a data set to lower dimensions for analysis, visualization or data compression.
- ... Hence, remove redundancy
- Change detection
- Classification

PCA dummy example: Linear regression

- Consider a normal linear regression model



- If we fit a normal linear regression we get:

```
coef(lm(y~x))
```

```
## (Intercept)           x
##    3.215334    1.969287
```

PCA dummy example: Linear regression

- Now consider this in terms of the empirical covariance matrix between x and y

```
V<-cov(cbind(x,y))
V
```

```
##           x        y
## x  8.418075 16.57760
## y 16.577604 36.42185
```

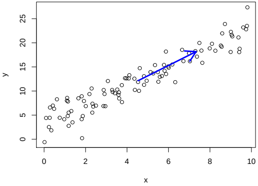
- The successive the most important orthogonal axis can be found by eigen-value decomposition
- For a positive-definite symmetric matrix we get: $V = UDU'$, where D is a diagonal matrix of eigen values, as in

```
e<-eigen(V)
e$vectors%*%diag(e$values)%*%t(e$vectors)
```

```
##           [,1]     [,2]
## [1,]  8.418075 16.57760
## [2,] 16.577604 36.42185
```

PCA dummy example: Linear regression

- The eigen vectors show the direction, and the eigen values the size (meassured in variance)
- Most important direction is:



- and that direction explains a big part of the overall variance:

```
e$values/sum(e$values)
```

```
## [1] 0.98393335 0.01606665
```

PCA dummy example: Linear regression

- Notice that the same analysis can be done with a build in function in R:

```
# Slope based on eigenvector
slope<-e$vectors[2,1]/e$vectors[1,1]
slope
```

```
## [1] 2.153593
```

```
coef(lm(y~x))[2]
```

```
##        x
## 1.969287
```

- The slopes are not exactly the same, why?

```
prcomp(cbind(x,y))
```

```
## Standard deviations (1, .., p=2):
## [1] 6.6422506 0.8487799
##
## Rotation (n x k) = (2 x 2):
##         PC1        PC2
## x 0.4211520  0.9069901
## y 0.9069901 -0.4211520
```

- except the standard deviations sqrt(eigen values) are listed instead of the variances

Exercise: Air quality in NYC

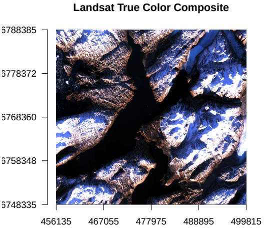
- The data set `airquality` contains 153 daily air quality measurements in New York, May to September 1973. Tha data set is build into R.
- Use only the first 4 columns of the data set, and figure out a way to exclude any rows which contains missing values.
- Then find the two main principal components.
- What fraction of the variance is explained by these two components?
- Which of the original data features is dominant in the first 2 components **Hint:** calculate the loads

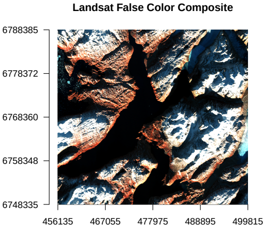
Working with raster data i R (Matlab, Python)

- Let us look at a Landsat 8 image using the visible bands and the NIR band (4 bands in total)

```
library(raster)
gr<-stack("/home/karina/teaching/ImageRS/scripts/data/greenland.tif")
gr.df <- as.data.frame(gr) # restructure data into a table (one band pr column)
 b2 <- gr[[1]] # an image with band 2
 b3 <- gr[[2]]
 b4 <- gr[[3]]
 b5 <- gr[[4]]

grRGB <- stack(b4, b3, b2) # true color composite
grFCC <- stack(b5, b4, b3) # false color composite
par(mfrow = c(1,2))
plotRGB(grRGB, axes=TRUE, stretch="lin", main="Landsat True Color Composite",zlim=c(0,0.5))
plotRGB(grFCC, axes=TRUE, stretch="lin", main="Landsat False Color Composite",zlim=c(0,0.5))
```



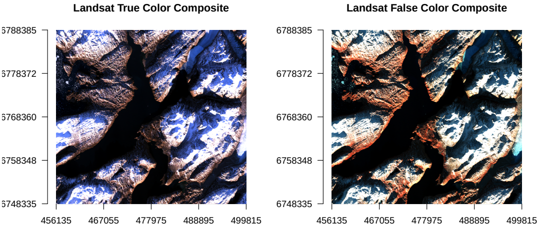


Exercise Landsat 8 image of Yukon River, Alaska, Part 1

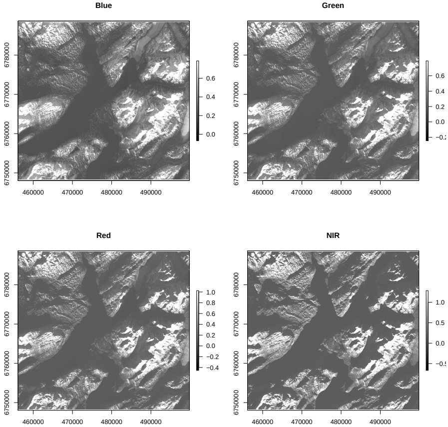
- Read the data file "yukon.tif" into R/matlab/Python **Hint**: see R/matlab/python scripts
- Produce gray scale plot of the individual bands
- Plot histograms of the bands (remember to reshape). Why are the histograms for the bands multimodal?
- Standardize data

PCA an example from Greenland

- Let us look at a Landsat 8 image using the visible bands and the NIR band (4 bands in total)

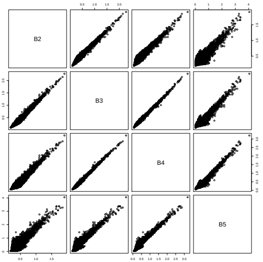


Plot of the individual bands



- Notice the high correlation between the bands

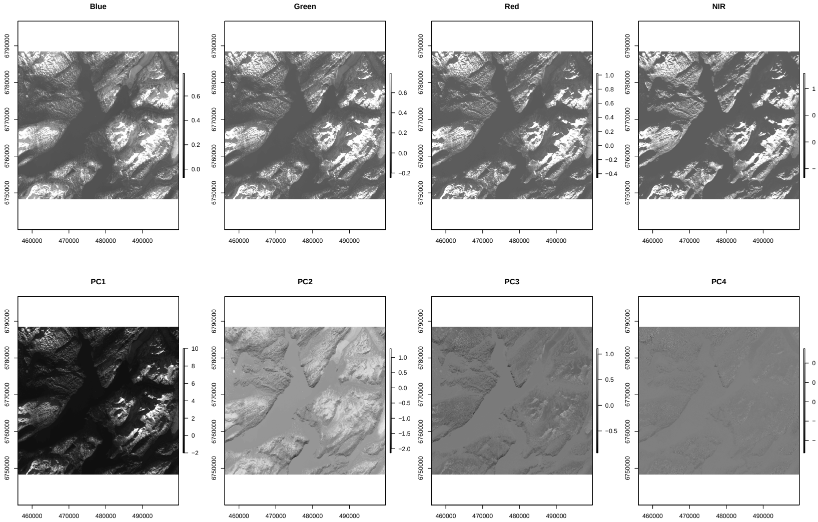
Correlation among the individual bands



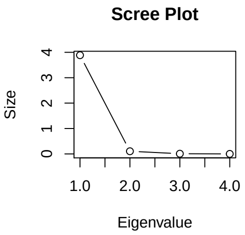

```r
# Correlation matrix
mycor
```

```
##           B2        B3        B4        B5
## B2 1.0000000 0.9795341 0.9547308 0.9004098
## B3 0.9795341 1.0000000 0.9934955 0.9600534
## B4 0.9547308 0.9934955 1.0000000 0.9825159
## B5 0.9004098 0.9600534 0.9825159 1.0000000
```

The principal components



Eigenvalues and eigenvectors




```r
# Percentage of variance
pca$sdev^2/sum(pca$sdev^2)*100
```

```
## [1] 97.14827357  2.63344946  0.19715462  0.02112234
```

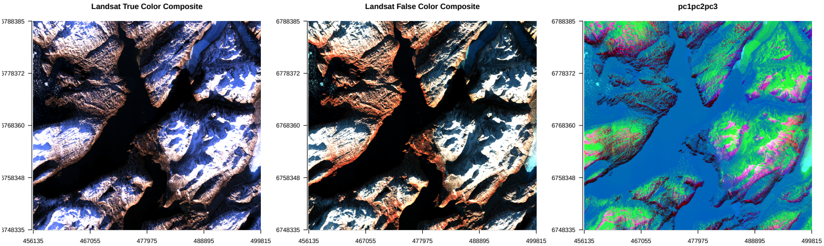
```r
#Eigenvectors
pca$rotation
```

```
##          PC1        PC2        PC3        PC4
## B2 0.4934082  0.7033641 -0.4810663  0.1743632
## B3 0.5061111  0.1496139  0.4851389 -0.6972140
## B4 0.5058198 -0.1794793  0.5021489  0.6780707
## B5 0.4945161 -0.6713284 -0.5301527 -0.1539810
```

A bit of everything

B2 and B5 dominate

False color composite of the first 3 PCs



- The PC color plot show enhanced detail in the image

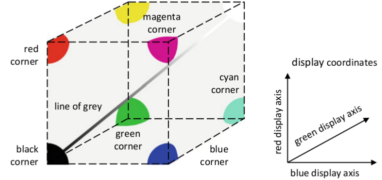


Fig. 6.11 The colour space in which three bands of remotely sensed image data are displayed; fully correlated data points would fall along the line of grey

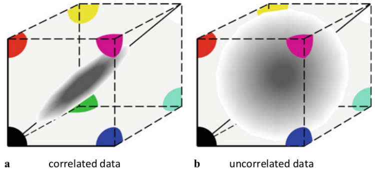


Figure 6: source: Remote Sensing Digital Image Analysis, John A. Richards

Example of change detection - wild fire

- PCA can be used to enhance change between two images (before and after an event)
- In the event of a wild fire we would expect changes in the NIR band in the area affected by the fire
- but no or little change in the remaining areas

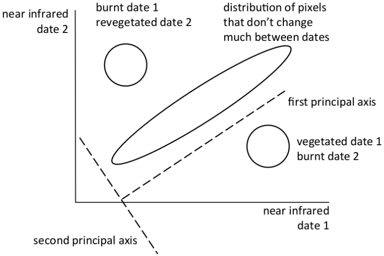


Figure 7: source: Remote Sensing Digital Image Analysis, John A. Richards

- In which principal components do we expect to see the change?

Exercise PCA of a Landsat 8 image of Yukon River, Alaska

- Calculate the covariance/correlation matrix
- Calculate principal components of the six spectral bands
- List singular or eigen values and corresponding eigenvectors and display the images of the transposed data (PC-data)
- Calculate the covariance matrix of PC scores and compare with eigenvalues.
- Create a false color composite of the first 3 PCs, what do you see?
- How much of the variance is described by the first 3 components?
- Plot histograms of the PC data, what do you see?
- Which band is PC2 dominated by, and how is this seen in the image of PC2. Hence, which surface type is mainly shown?

Exercise of change detection - wild fire

- Perform a PCA using the two Landsat 8 images one before and after a fire. The two images are called "L8pre.tif" and "L8burn.tif". **Hint:** you need to combine the 4 bands from each image to one data set.
- Which PC detects the change?
- Make a false color composite of the 3 PCs that best depicts the change

PCA on a self selected image L8/S2

- Select a sub region of an image and export via GEE.
- Perform a PCA analysis
- Present what you see

Report

- Summarize the results of the exercises
- Write a small report/journal (3-4 pages) including figures
- Hence, you do not need to describe all exercises in detail, only the main results
- Attach code as an appendix