

Principal component analysis

Dai volti agli spettri di galassie

PCA for beginners

SCOPO:

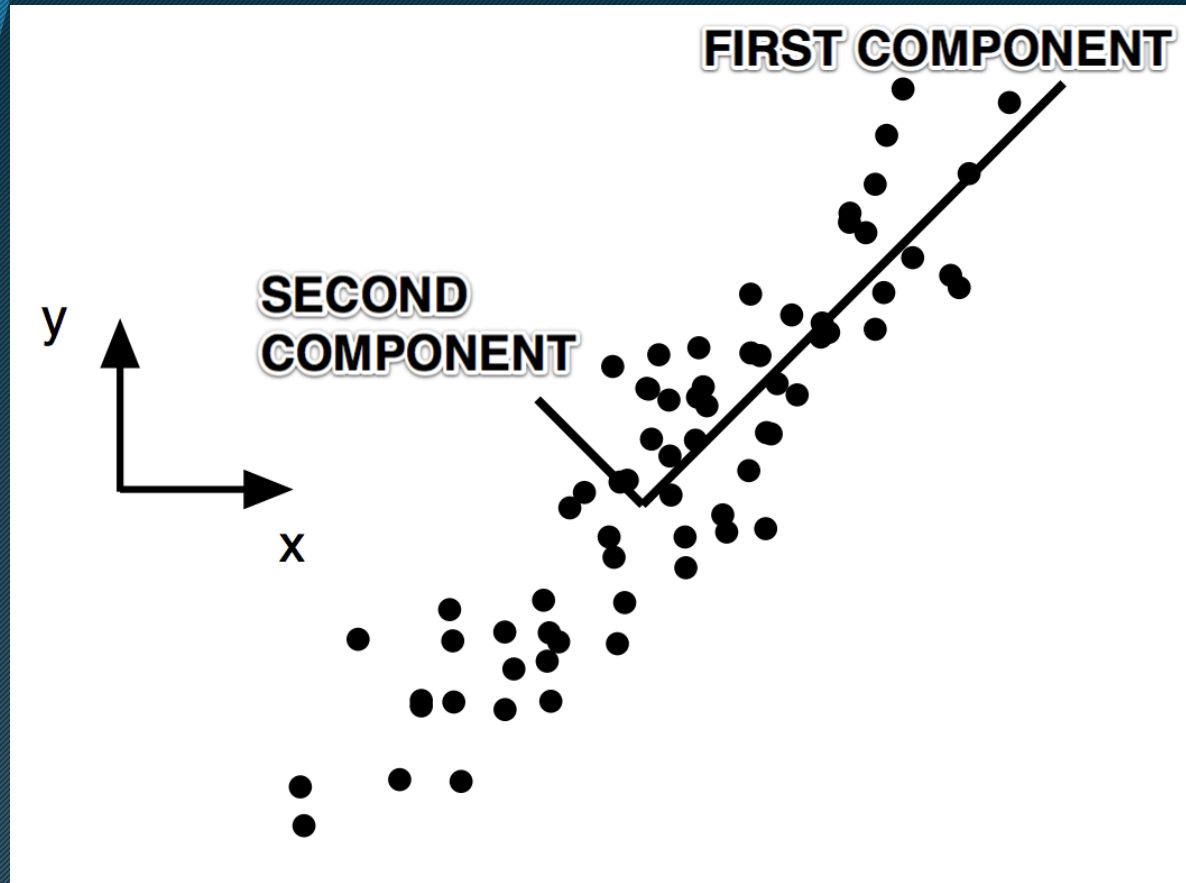
Ridurre il numero di variabili all'interno di una collezione di oggetti per descriverli piu' facilmente, sulla base delle caratteristiche ricorrenti.

Caso bidimensionale:

	x	y
	2.5	2.4
	0.5	0.7
	2.2	2.9
	1.9	2.2
Data =	3.1	3.0
	2.3	2.7
	2	1.6
	1	1.1
	1.5	1.6
	1.1	0.9

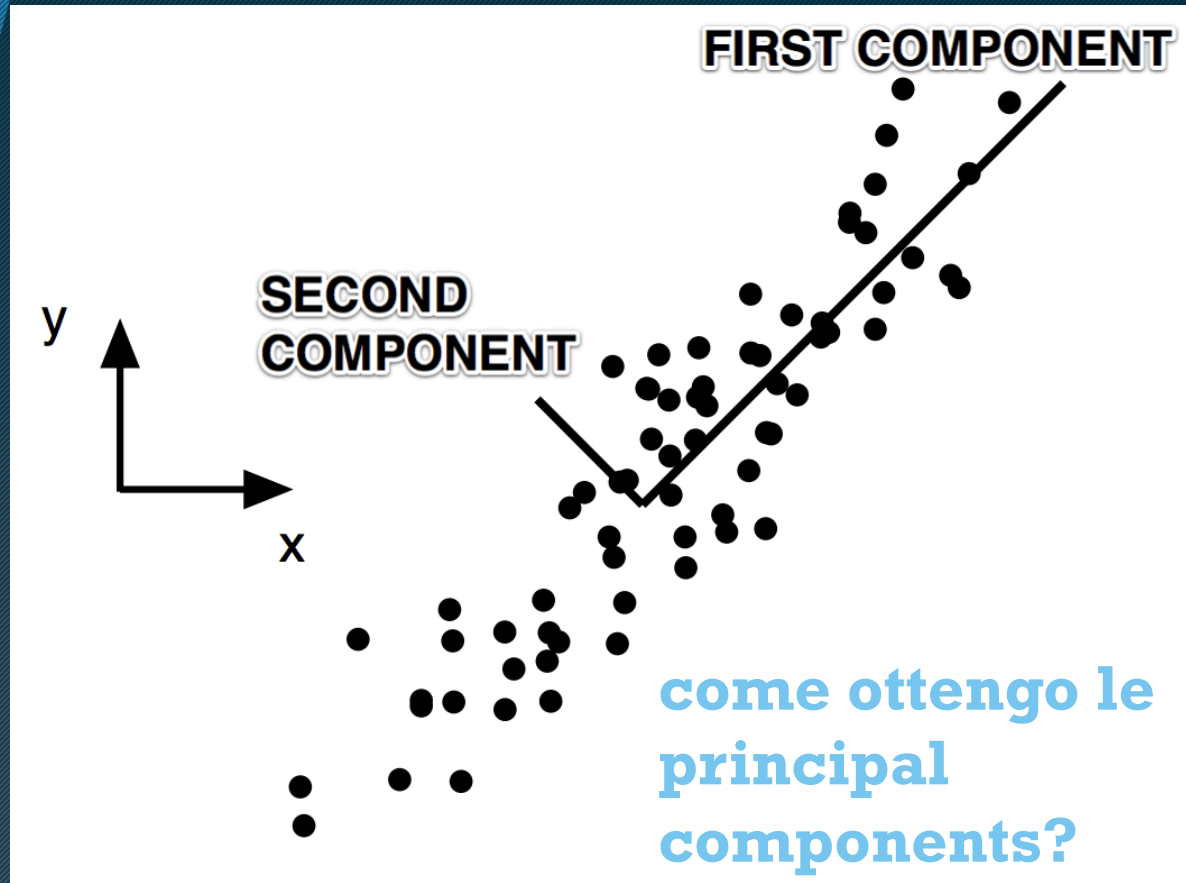
Caso bidimensionale:

PCA for beginners



PCA for beginners

Caso bidimensionale:



Caso bidimensionale:

1. sottraggo la media

	x	y
	.69	.49
	-1.31	-1.21
	.39	.99
	.09	.29
DataAdjust =	1.29	1.09
	.49	.79
	.19	-.31
	-.81	-.81
	-.31	-.31
	-.71	-1.01

Caso bidimensionale:

2. Calcolo la matrice di covarianza

$$ccv(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n - 1)}$$

3. Ricavo gli **autovettori** e gli **autovalori**. La prima componente e' l'autovettore legato all'autovalore piu' alto

Caso N-dimensionale:

Ora posso proiettare i miei punti sulla nuova base, formata dai 2 autovettori.

Ci sono applicazioni N-dimensionali, in cui le prime 3 o 4 componenti sono molto piu' significative delle altre N-4.

Caso N-dimensionale:

Posso descrivere tutto il set N-dimensionale avvalendomi solo di poche componenti.

Es. Ogni “oggetto” nel campione potrà essere descritto come combinazione lineare di pochi autovettori, tramite opportuni coefficienti.

UN PAIO DI APPLICAZIONI...

1. Riconoscimento facciale

COMPUTER VISION

I volti hanno in comune molte caratteristiche (2 occhi, naso, bocca...)

Tramite la PCA si possono ottenere degli autovettori per un insieme di volti, chiamati *eigenfaces*

I volti devono essere fotografati con le stesse condizioni di luce e devono essere posizionati con le caratteristiche fisiognomiche allineate il piu' possibile tra loro.

Ogni riga della mia matrice rappresentera' un'immagine

La matrice di covarianza avra' la dimensione della singola immagine, e cosi' gli autovettori

COMPUTER VISION

I volti hanno

Tramite la PCA
volti, chiamati

I volti devono
devono essere
il piu' possibile



(a...)

me di

lineate

Ogni riga della mia matrice rappresentera' un'immagine

La matrice di covarianza avra' la dimensione della singola immagine, e
cosi' gli autovettori

COMPUTER VISION

eigenfaces



Ognuno dei volti del campione potrà' essere descritto come
es. $75\%eigenface1 + 5\%eigenface2 + 10\%eigenface3 + \dots$
E ricostruito tramite un' opportuna combinazione

COMPUTER VISION



Se uso poche autofacce, posso ricostruire uno dei volti di partenza in modo sempre piu' preciso all'aumentare delle immagini di partenza

COMPUTER VISION



Se uso poche autofacce, posso ricostruire uno dei volti di partenza in modo sempre piu' preciso all'aumentare delle immagini di partenza

COMPUTER VISION



Se uso poche autofacce, posso ricostruire uno dei volti di partenza in modo sempre piu' preciso all'aumentare delle immagini di partenza

COMPUTER VISION



Se uso poche autofacce, posso ricostruire uno dei volti di partenza in modo sempre piu' preciso all'aumentare delle immagini di partenza

COMPUTER VISION: face recognition

Tramite le “autofacce” si puo’ far riconoscere una fisionomia a una macchina

CREAZIONE DEL DATABASE

1. Creo un database di foto

COMPUTER VISION: face recognition

Tramite le “autofacce” si puo’ far riconoscere una fisionomia a una macchina

CREAZIONE DEL DATABASE

1. Creo un database di foto
2. Calcolo le autofacce principali

COMPUTER VISION: face recognition

Tramite le
a una macchina

CREAZIONE

1. Creo un

2. Calcolo

```
clear all
close all
load yalefaces
[h,w,n] = size(yalefaces);
d = h*w;
% vectorize images
x = reshape(yalefaces,[d n]);
x = double(x);
%subtract mean
x=bsxfun(@minus, x', mean(x'))';
% calculate covariance
s = cov(x');
% obtain eigenvalue & eigenvector
[V,D] = eig(s);
eigval = diag(D);
% sort eigenvalues in descending order
eigval = eigval(end:-1:1);
V = fliplr(V);
% show 0th through 15th principal eigenvectors
eig0 = reshape(mean(x,2), [h,w]);
figure,subplot(4,4,1)
imagesc(eig0)
colormap gray
for i = 1:15
subplot(4,4,i+1)
imagesc(reshape(V(:,i),h,w))
end
```

fisionomia

COMPUTER VISION: face recognition

Tramite le “autofacce” si puo’ far riconoscere una fisionomia a una macchina

CREAZIONE DEL DATABASE

```
% evaluate the number of principal components needed to represent 95% Total variance.
eigsum = sum(eigval);
csum = 0;
for i = 1:d
csum = csum + eigval(i);
tv = csum/eigsum;
if tv > 0.95
k95 = i;
break
end
end
```

COMPUTER VISION: face recognition

Tramite le “autofacce” si puo' far riconoscere una fisionomia a una macchina

CREAZIONE DEL DATABASE

1. Creo un database di foto
2. Calcolo le autofacce principali
3. Calcolo i coefficienti della combinazione lineare per ogni foto nel database

COMPUTER VISION: face recognition

Tramite le “autofacce” si puo' far riconoscere una fisionomia a una macchina

RICONOSCIMENTO

1. Modifico la foto da riconoscere
2. Ne calcolo i coefficienti sulla base delle autofacce
3. Confronto i coefficienti ottenuti con quelli registrati nel database

UN PAIO DI APPLICAZIONI...

1. Riconoscimento facciale

UN PAIO DI APPLICAZIONI...

1. Riconoscimento facciale
2. Astrofisica

ASTROFISICA: gli spettri

Gli spettri elettromagnetici delle galassie (o delle stelle) hanno diverse caratteristiche ma sono accomunati dallo stesso processo generativo.

Dagli anni '50 la PCA si applica con successo a

1. classificazione di spettri stellari o galattici
2. ricostruzione di porzioni di spettro danneggiate dal rumore o assenti per questioni di redshift
3. pulizia statistica degli spettri dal rumore

ASTROFISICA: gli spettri

Spettri in una survey portati a rest frame e normalizzati

Matrice dei dati: spettri sulle righe

$$\mathbf{M} = \begin{bmatrix} f_{1\lambda_1} & \cdots & f_{1\lambda_n} \\ \vdots & \ddots & \vdots \\ f_{N\lambda_1} & \cdots & f_{N\lambda_n} \end{bmatrix}$$

Ogni spettro rappresenta una diversa realizzazione per le diverse lunghezze d'onda in un range prestabilito

ASTROFISICA: gli spettri

Matrice di covarianza

$$C_{\lambda_1, \lambda_2} = \frac{1}{n-1} \sum_{i=1}^n f_{\lambda_1}^i f_{\lambda_2}^i$$

Autovettori o *autospettri*

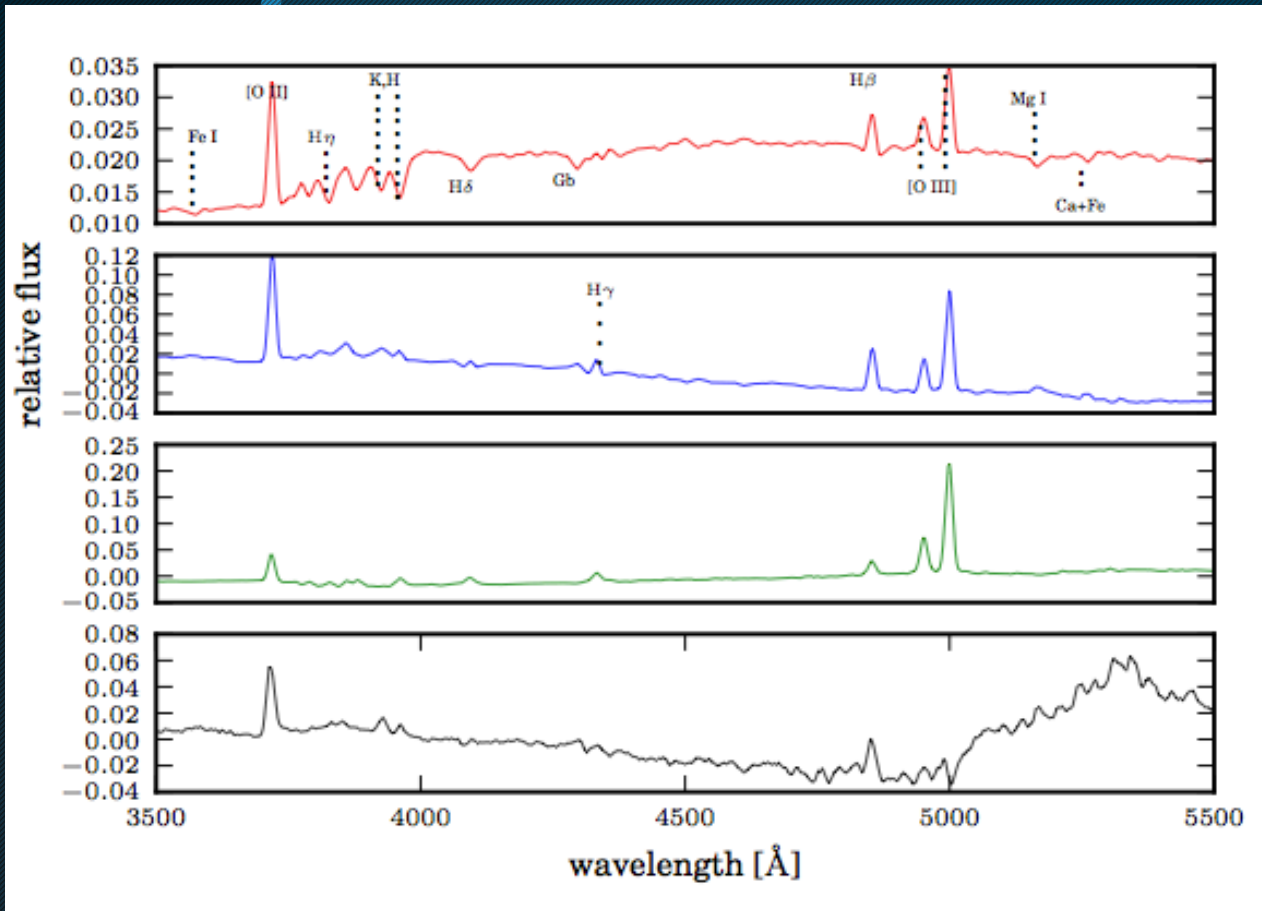
$$C_{\lambda_1, \lambda_2} = \sum_{i=1}^n e_{\lambda_1}^i \Lambda_i e_{\lambda_2}^i$$

Con cui posso descrivere “tutta” la survey. Quanti sono? Me ne bastano pochi?

ASTROFISICA: gli Autospettri



VIPERS: redshift survey appena conclusa a ESO VLT, circa 100000 spettri a $0.5 < z < 1$. (www.vipers.inaf.it)

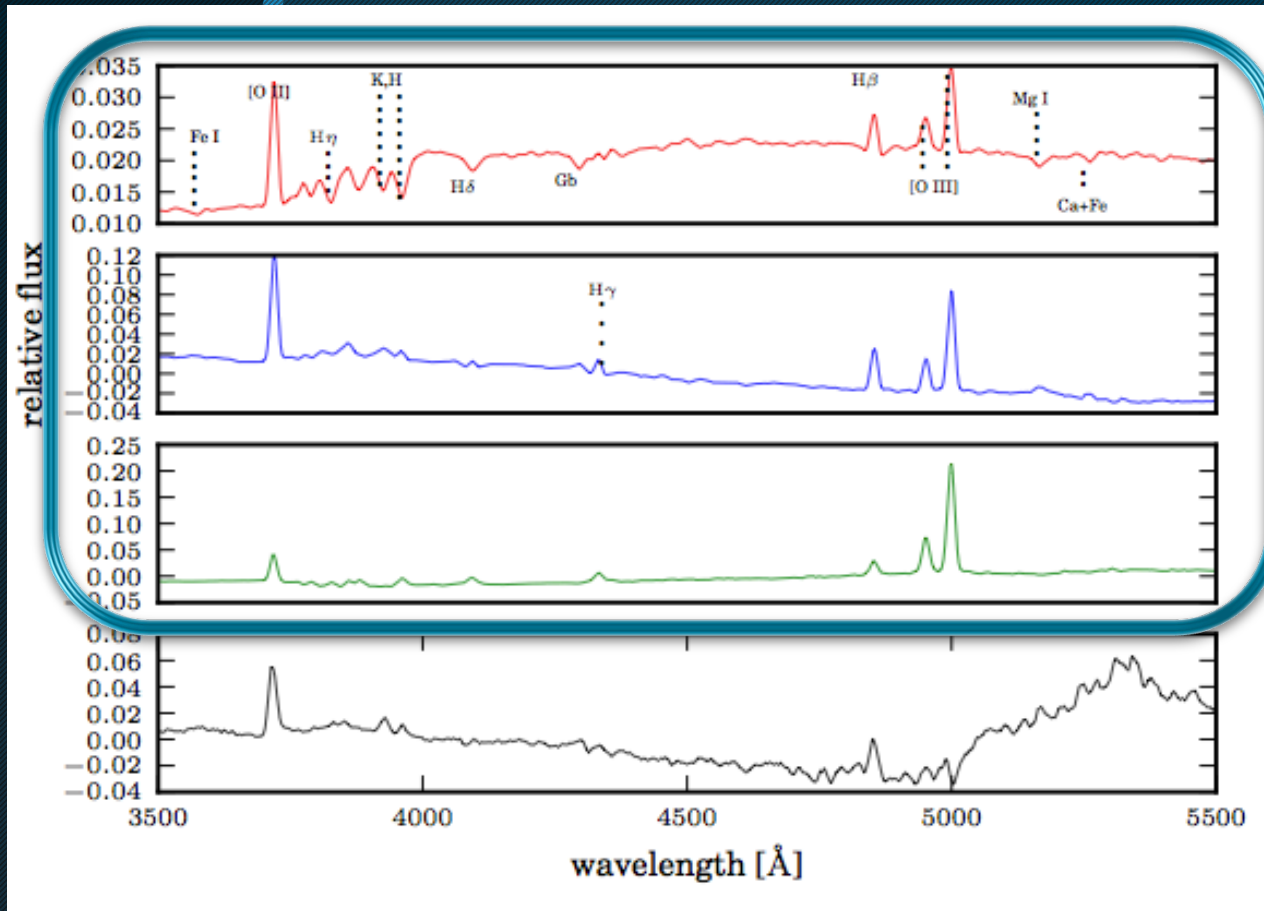


eigen spectra

ASTROFISICA: gli Autospettri



VIPERS: redshift survey appena conclusa a ESO VLT, circa 100000 spettri a $0.5 < z < 1$. (www.vipers.inaf.it)



eigen spectra

3 autospettri !!!!

90% della survey

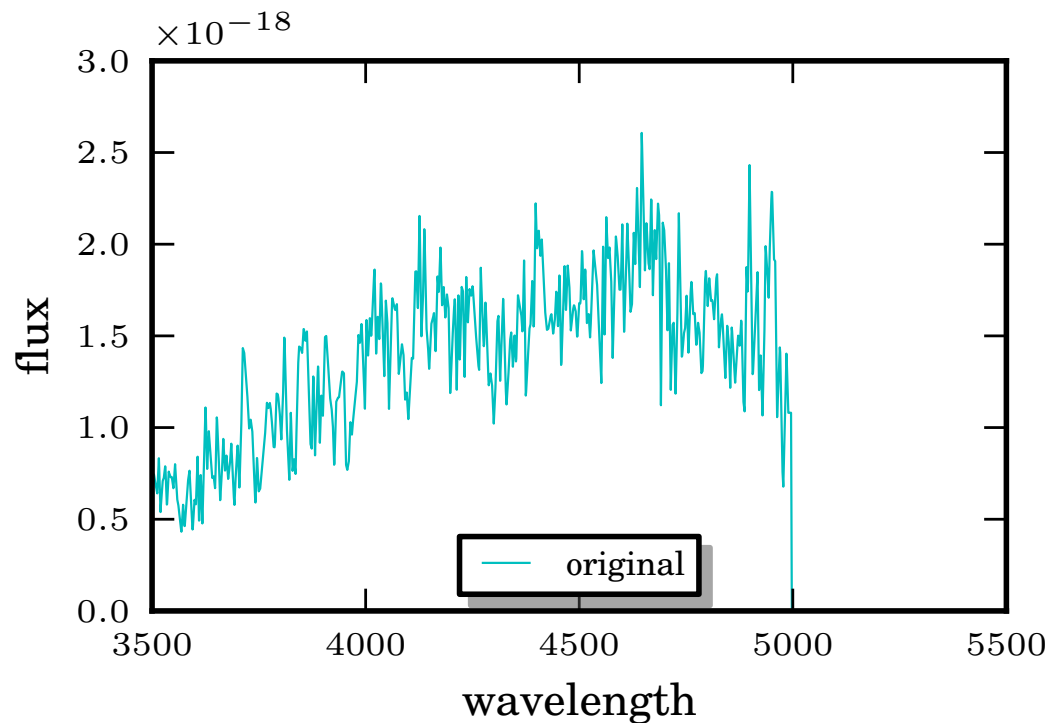
No rumore

ASTROFISICA: gli Autospettri



Ogni spettro puo'essere descritto tramite 3 coefficienti
Posso usare questa combinazione lineare per:

1.RIPARARE GLI SPETTRI DOVE NON HO SEGNALE

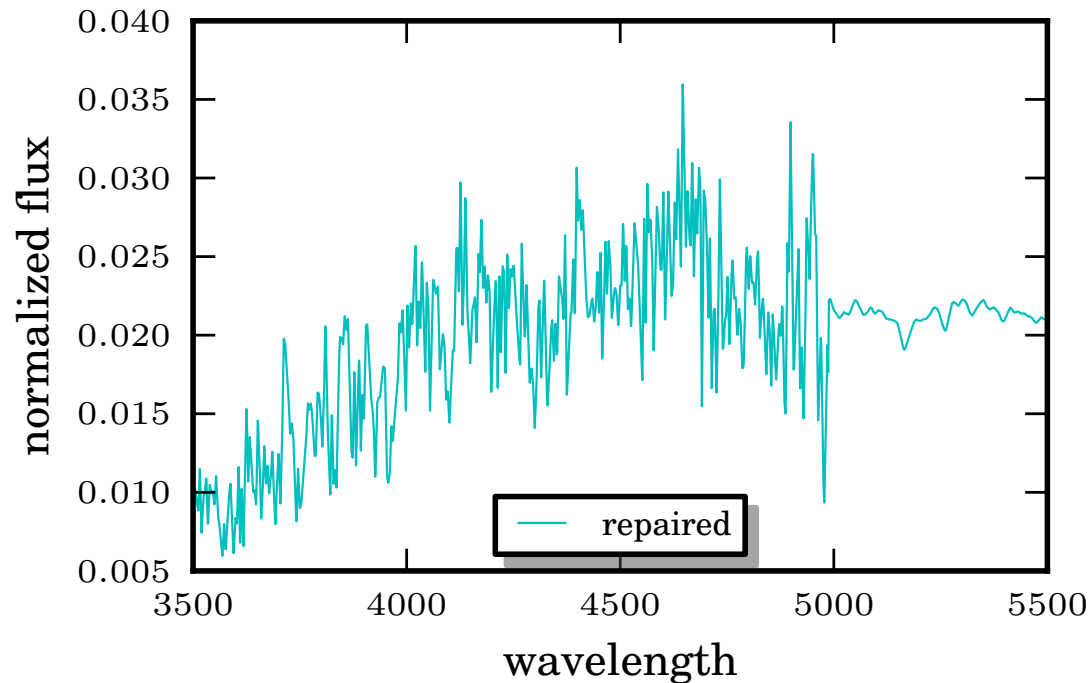


ASTROFISICA: gli Autospettri



Ogni spettro puo'essere descritto tramite 3 coefficienti
Posso usare questa combinazione lineare per:

1. RIPARARE GLI SPETTRI DOVE NON HO SEGNALE

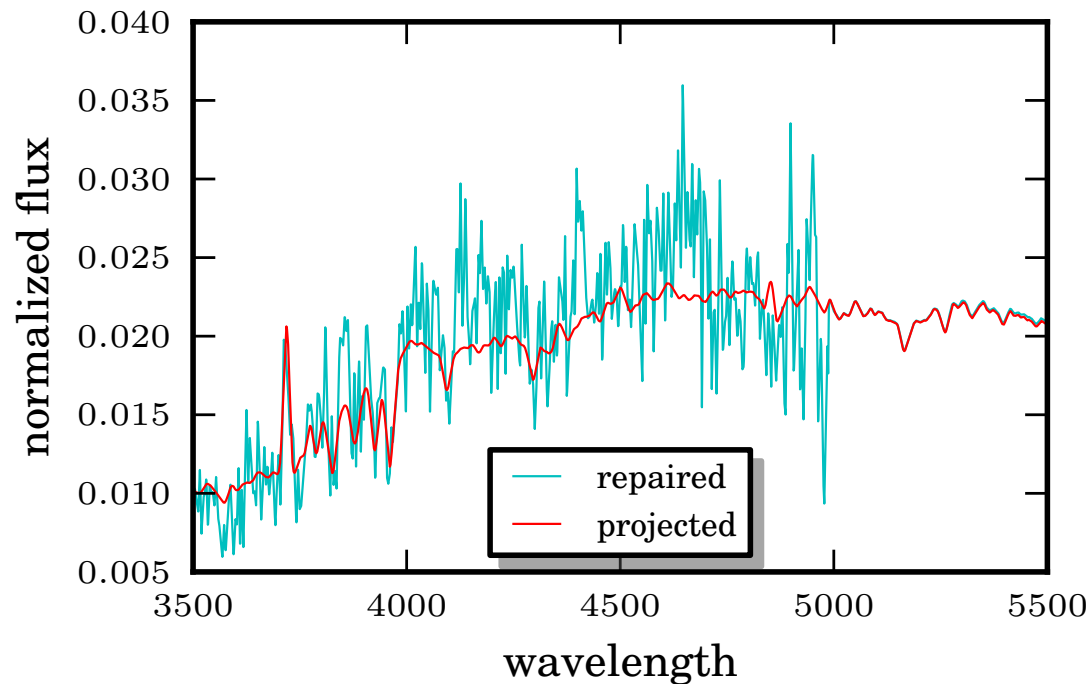


ASTROFISICA: gli Autospettri



Ogni spettro puo'essere descritto tramite 3 coefficienti
Posso usare questa combinazione lineare per:

2.RIPULIRE STATISTICAMENTE GLI SPETTRI DAL RUMORE

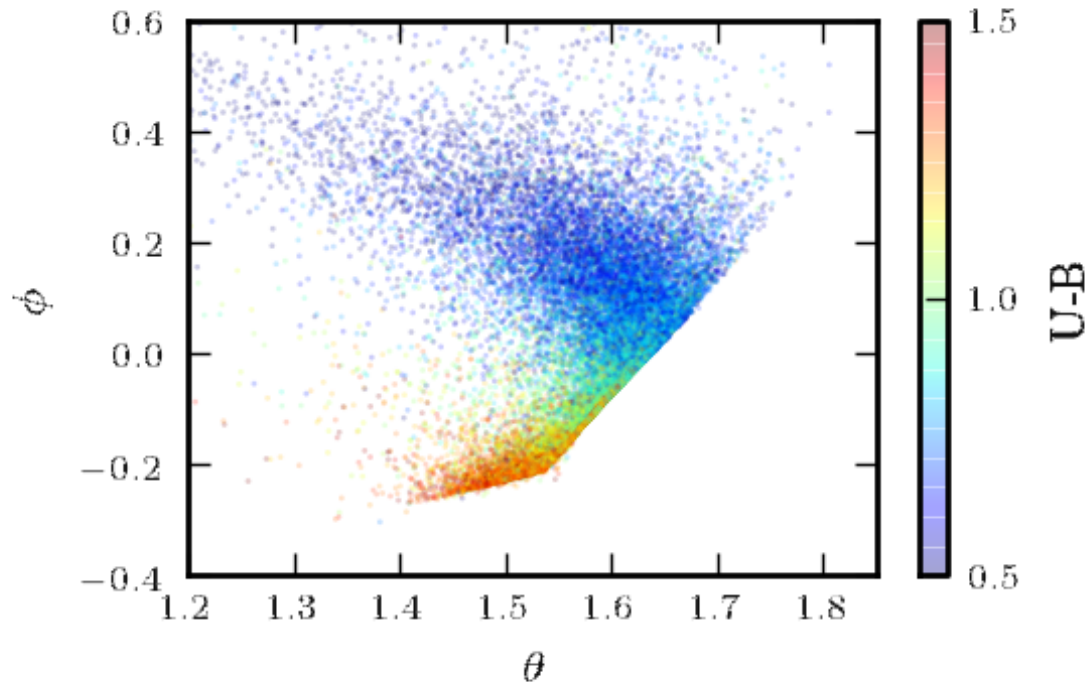


ASTROFISICA: gli Autospettri



Ogni spettro puo'essere descritto tramite 3 coefficienti
Posso usare questa combinazione lineare per:

3.CLASSIFICARE GLI SPETTRI



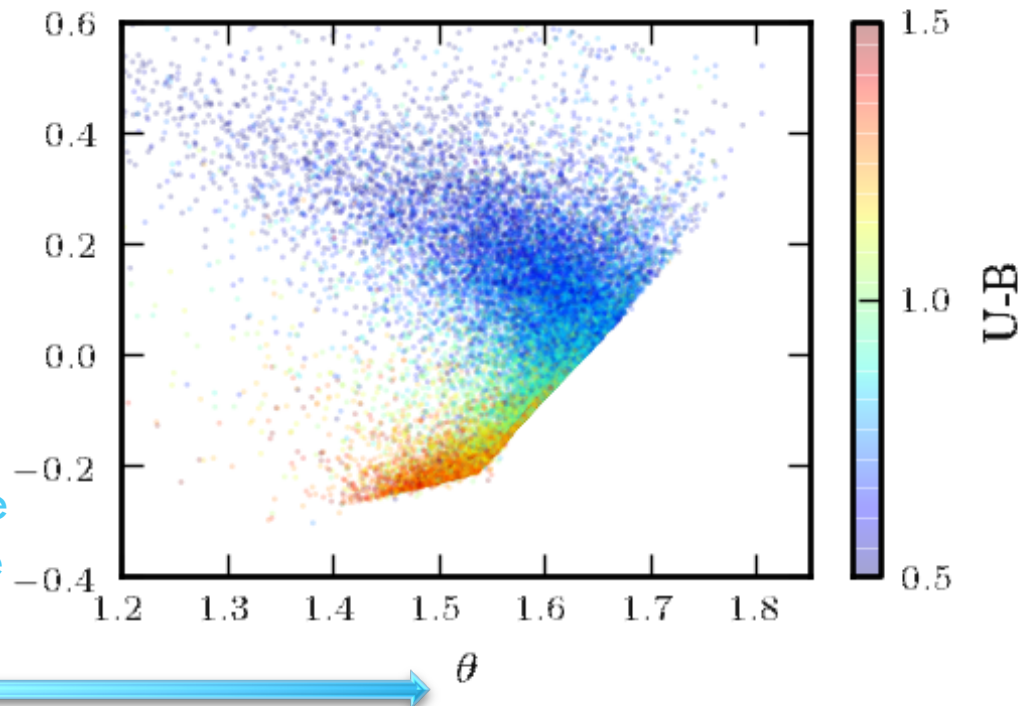
ASTROFISICA: gli Autospettri



Ogni spettro puo'essere descritto tramite 3 coefficienti
Posso usare questa combinazione lineare per:

3. CLASSIFICARE GLI SPETTRI

→ ϕ



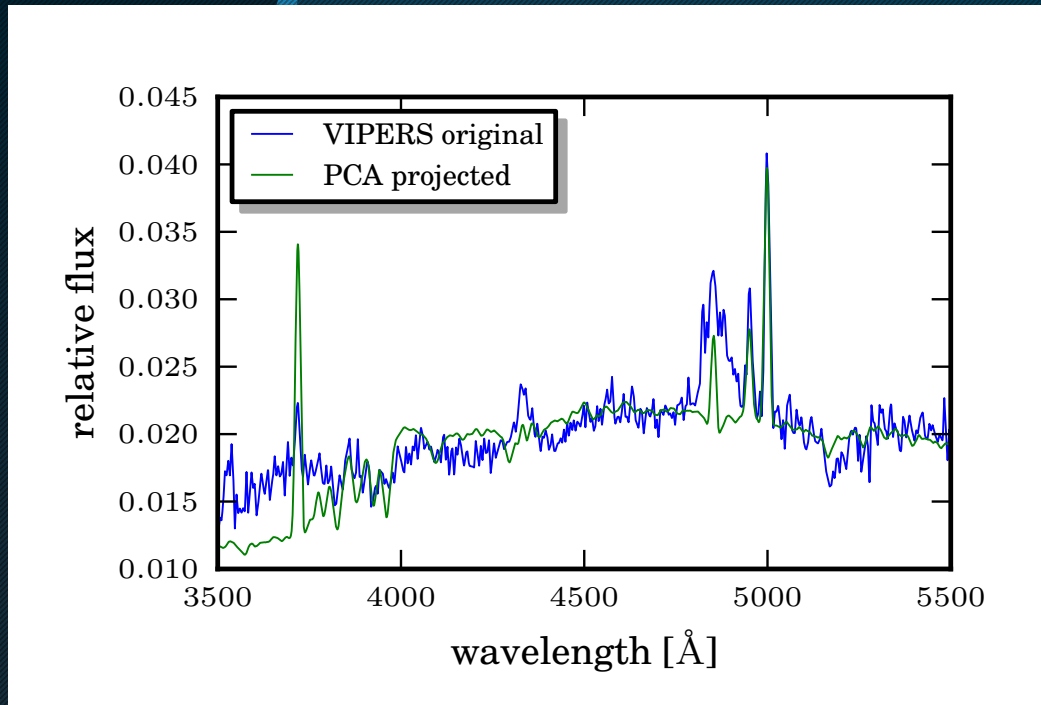
→ θ

Coefficienti della
Combinazione lineare
in coordinate sferiche

ASTROFISICA: gli Autospettri...nota bene



Non sempre la fase di “pulizia” rispecchia la forma piu’ probabile dello spettro senza rumore, es. AGN

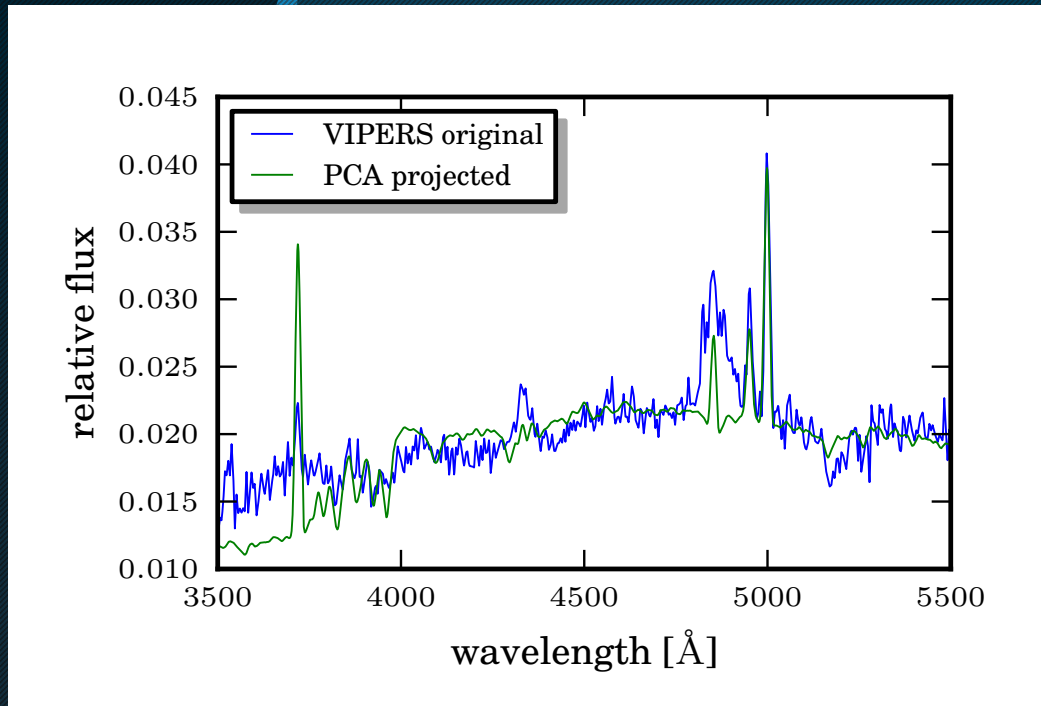


I primi autospettri non contengo info sugli spettri “rari”.

ASTROFISICA: gli Autospettri...nota bene



Non sempre la fase di “pulizia” rispecchia la forma piu’ probabile dello spettro senza rumore, es. AGN



I primi autospettri non contengono info sugli spettri “rari”.
Sulla base del χ^2 del cleaning posso selezionare automaticamente gli AGN!!

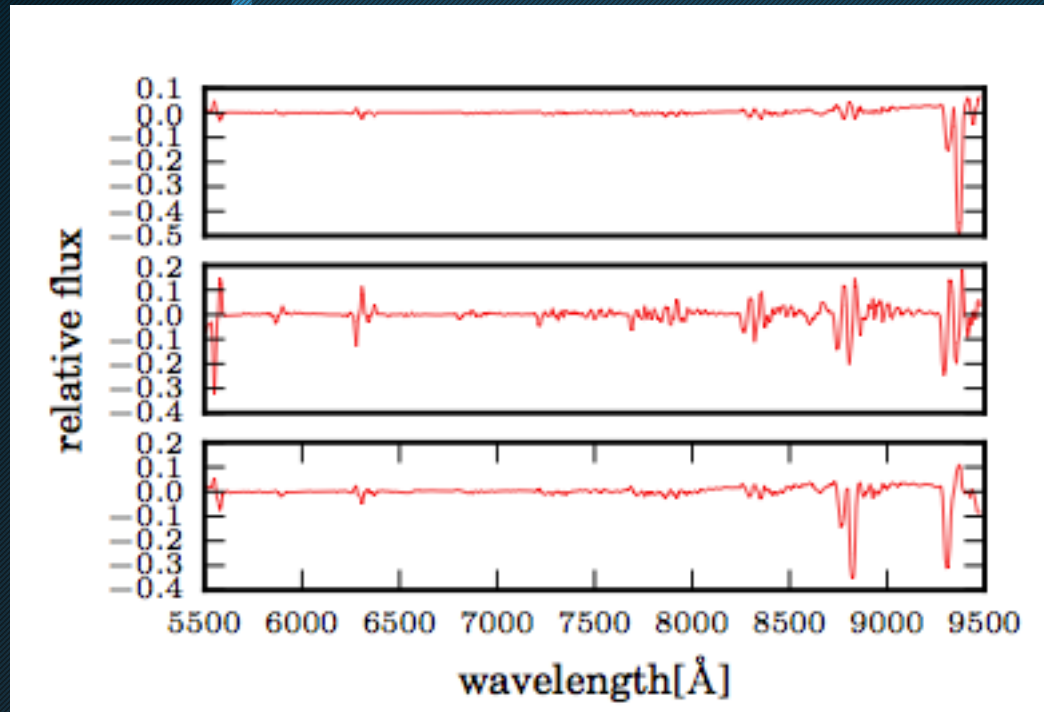
ASTROFISICA: gli Autospettri di cielo



Se invece creiamo la matrice dei dati lasciando gli spettri in observed frame:

1. il segnale dello spettro sarà statisticamente “smorzato”
2. la componente di cielo sarà dominante
3. gli autospettri rappresenteranno la componente di cielo

Sky eigenspectra



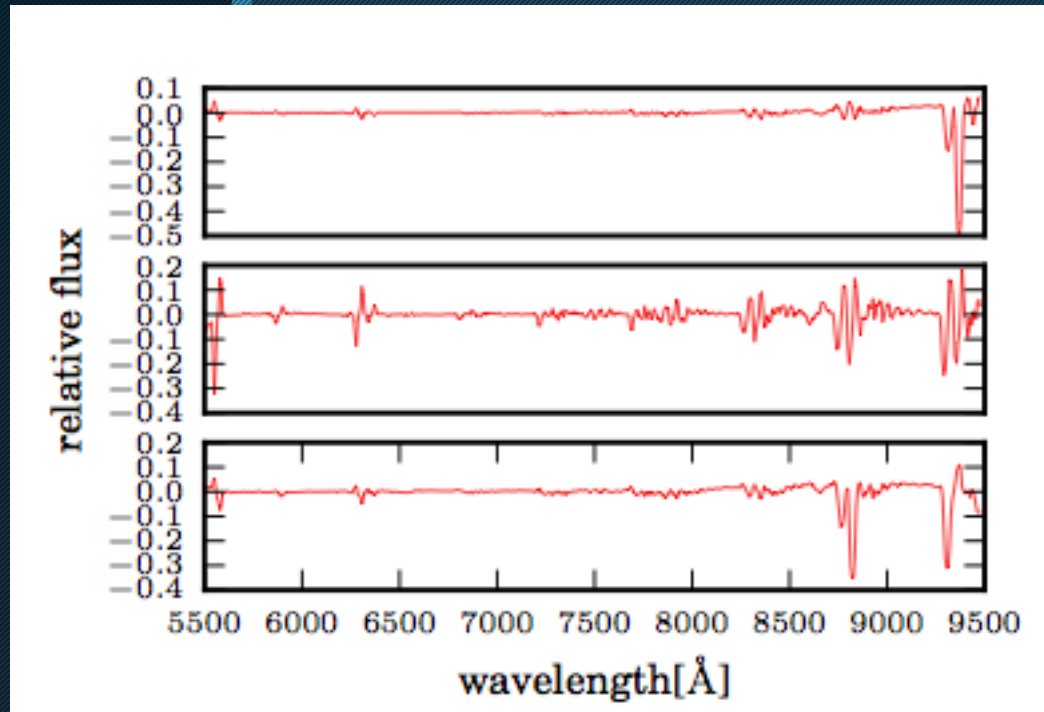
ASTROFISICA: gli Autospettri di cielo



Se invece creiamo la matrice dei dati lasciando gli spettri in observed frame:

1. il segnale dello spettro sarà statisticamente “smorzato”
2. la componente di cielo sarà dominante
3. gli autospettri rappresenteranno la componente di cielo

Sky eigenspectra



Anche in questo caso ne bastano 3!!

UN PAIO DI APPLICAZIONI...

1. Riconoscimento facciale
2. Astrofisica

UN PAIO DI APPLICAZIONI...

1. Riconoscimento facciale
2. Astrofisica
3. Altre applicazioni

ALTRE APPLICAZIONI ...!!

Neuroscienza

Biologia

Farmacia

Agricoltura

Ecologia

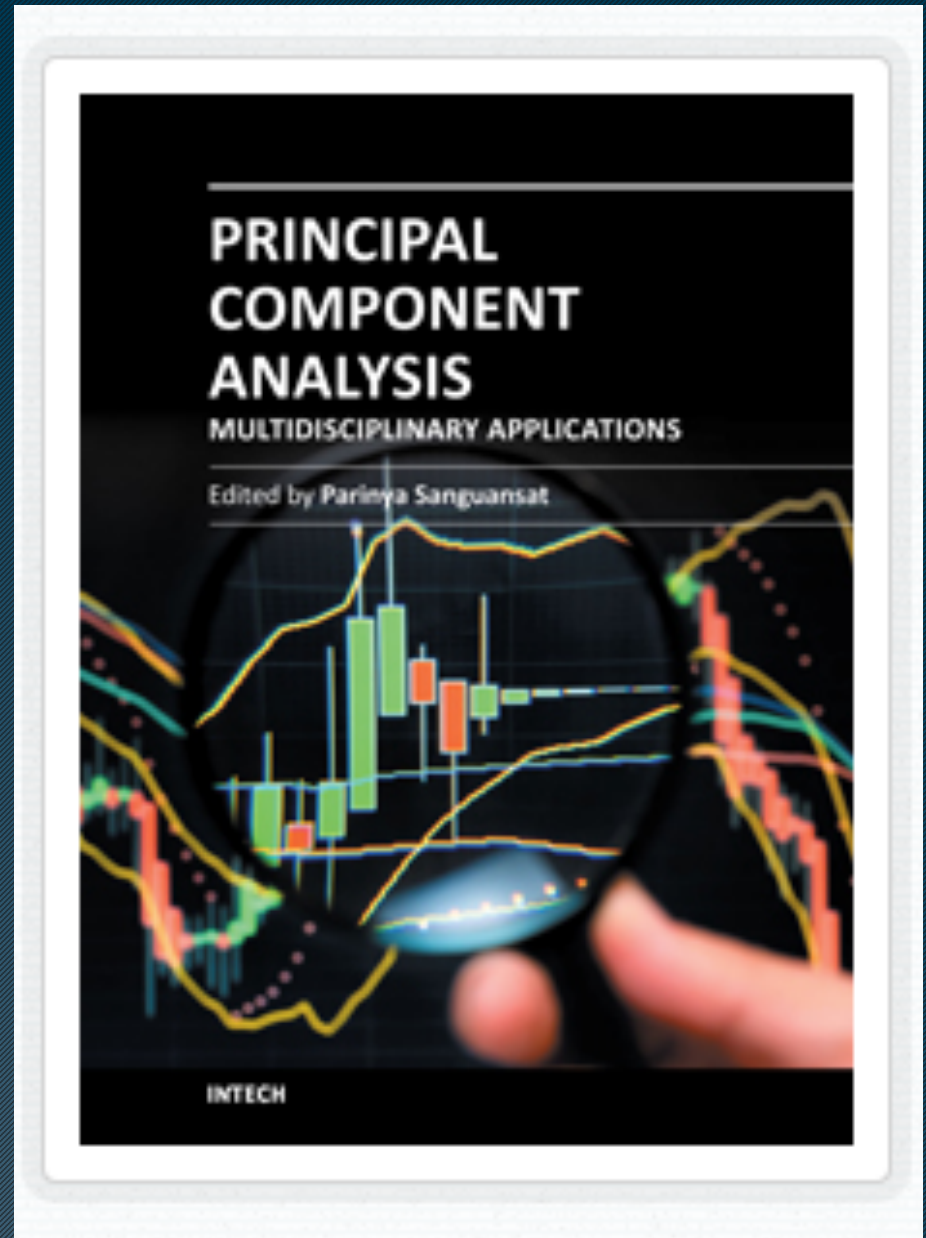
Salute

Architettura

Finanza

Taxonomia

...



ALTRE APPLICAZIONI ...!!

Neuroscienza

Biologia

Farmacia

Agricoltura

Ecologia

Salute

Architettura

Finanza

Taxonomia

...

GRAZIE!

