Subject: Re: Principal component analysis
Posted by David Fanning on Wed, 05 Dec 2007 14:30:01 GMT
View Forum Message <> Reply to Message

Haje Korth writes:

> I am puzzled by principal component analysis. I calculated the eigenvalues
> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten me
> why the results are completely different? I have tried different keywords to
> see whether I can match them by trial and error, but I had no success. There
> must be someone out there who undertstands this much better than I do.

Did you read this article:

http://www.dfanning.com/code_tips/pca.html

I don't know what IMSP_PRINC_COMP is, but the article
will maybe help you come to terms with how different
PCA approaches might be different.

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by Haje Korth on Wed, 05 Dec 2007 14:36:18 GMT
View Forum Message <> Reply to Message

David,
I read your article three times to not embarass myself here. (I am omitting
the smiley here) IMSL_PRINC_COMP is part of the IDL Analyst library. I dug a
little further and this is what I found the IMSL routine gives the same
eigenvectors as obtained form EIGENQL(a). However, in PCOMP the call to get
the eigenvalues is EIGENQL(CORRELATE(a)). So which set is the correct one to
use?

Haje

"David Fanning" <news@dfanning.com> wrote in message
news:MPG.21c063d562d85c4a98a126@news.frii.com...

> Haje Korth writes:
>
>> I am puzzled by principal component analysis. I calculated the
>> eigenvalues
>> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten me
>> why the results are completely different? I have tried different keywords
>> to
>> see whether I can match them by trial and error, but I had no success.
>> There
>> must be someone out there who undertstands this much better than I do.
>
> Did you read this article:
>
> http://www.dfanning.com/code_tips/pca.html
>
> I don't know what IMSP_PRINC_COMP is, but the article
> will maybe help you come to terms with how different
> PCA approaches might be different.
>
> Cheers,
>
> David
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming: http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by David Fanning on Wed, 05 Dec 2007 14:49:32 GMT
View Forum Message <> Reply to Message

Haje Korth writes:

> I read your article three times to not embarass myself here. (I am omitting
> the smiley here) IMSL_PRINC_COMP is part of the IDL Analyst library. I dug a
> little further and this is what I found the IMSL routine gives the same
> eigenvectors as obtained form EIGENQL(a). However, in PCOMP the call to get
> the eigenvalues is EIGENQL(CORRELATE(a)). So which set is the correct one to
> use?

Don't know. You will have to ask someone whose last math
class wasn't over 30 years ago. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by David Fanning on Wed, 05 Dec 2007 14:57:18 GMT
View Forum Message <> Reply to Message

David Fanning writes:

> Don't know. You will have to ask someone whose last math
> class wasn't over 30 years ago. :-)

I will say, though, that when I was looking into this
I noticed that what you pass in really doesn't matter
as long as you preserve the "vector" nature of the data.
That is to say, the absolute magnitude of the vectors
that comes out is irrelevant. The *relative* magnitudes
of the orthogonal vectors "explains" the data. So I would
say the various ways you can manipulate the data prior to
sending it to analysis are "scaling" functions that don't
fundamentally change the nature of the result.

At least that is my hand-waving explanation. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by Vince Hradil on Wed, 05 Dec 2007 15:12:09 GMT
View Forum Message <> Reply to Message

On Dec 5, 8:00 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
> Hi,
> I am puzzled by principal component analysis. I calculated the eigenvalues

> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten me
> why the results are completely different? I have tried different keywords to
> see whether I can match them by trial and error, but I had no success. There
> must be someone out there who undertstands this much better than I do.
>
> Thanks so much,
> Haje
>
> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
> 2.24227 0.757732 0.000000
> IDL> ev=imsl_princ_comp(a) & print,ev
> 9.53359 -5.19751 2.66392

From the HELP:

Syntax
Result = IMSL_PRINC_COMP(covariances [, /COV_MATRIX]
[, /CORR_MATRIX] [, CORRELATIONS=variable] [, CUM_PERCENT=variable] [,
DF=variable] [, /DOUBLE] [, EIGENVECTORS=variable] [,
STDEV=variable])

Note that IMSL_PRINC_COMP requires that you pass the covariance or
correlation matrix - not the vectors.

Subject: Re: Principal component analysis
Posted by Vince Hradil on Wed, 05 Dec 2007 15:14:14 GMT
View Forum Message <> Reply to Message

On Dec 5, 9:12 am, Vince Hradil <hrad...@yahoo.com> wrote:
> On Dec 5, 8:00 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
>
>> Hi,
>> I am puzzled by principal component analysis. I calculated the eigenvalues
>> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten me
>> why the results are completely different? I have tried different keywords to
>> see whether I can match them by trial and error, but I had no success. There
>> must be someone out there who undertstands this much better than I do.
>
>> Thanks so much,
>> Haje
>
>> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
>> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
>> 2.24227 0.757732 0.000000
>> IDL> ev=imsl_princ_comp(a) & print,ev
>> 9.53359 -5.19751 2.66392

>
> From the HELP:
>
> Syntax
> Result = IMSL_PRINC_COMP(covariances [, /COV_MATRIX]
> [, /CORR_MATRIX] [, CORRELATIONS=variable] [, CUM_PERCENT=variable] [,
> DF=variable] [, /DOUBLE] [, EIGENVECTORS=variable] [,
> STDEV=variable])
>
> Note that IMSL_PRINC_COMP requires that you pass the covariance or
> correlation matrix - not the vectors.

so maybe try
ev=imsl_princ_comp(correlate(a,/covariance) & print, ev
(I don't have an analyst license)

Subject: Re: Principal component analysis
Posted by Mort Canty on Wed, 05 Dec 2007 15:24:33 GMT
View Forum Message <> Reply to Message

Haje Korth schrieb:
> Hi,
> I am puzzled by principal component analysis. I calculated the eigenvalues
> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten me
> why the results are completely different? I have tried different keywords to
> see whether I can match them by trial and error, but I had no success. There
> must be someone out there who undertstands this much better than I do.
>
> Thanks so much,
> Haje
>
>
> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
> 2.24227 0.757732 0.000000
> IDL> ev=imsl_princ_comp(a) & print,ev
> 9.53359 -5.19751 2.66392
>
>

Haven't the foggiest what imsl_princ_comp() does, but pcomp() is correct:

pro pca
a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
; covariance matrix
s1 = correlate(a,/covariance)
print, s1, ' '

```
; correlation matrix
s2 = correlate(a)
print, s2, ' '
; diagonalize
print, eigenql(s1)
print, eigenql(s2), ' '
; compare
pca=pcomp(a,eigenvalues=ev,/covariance) & print,transpose(ev)
pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
end

      12.3333      2.33333     -19.6667
      2.33333      4.33333     -5.66667
     -19.6667     -5.66667      32.3333

      1.00000     0.319173    -0.984839
     0.319173      1.00000    -0.478731
    -0.984839    -0.478731      1.00000

      45.2906      3.70938-1.52795e-006
      2.24227     0.757732-5.49480e-008

      45.2906      3.70938-1.52795e-006
      2.24227     0.757732     0.000000
```

Mort

Subject: Re: Principal component analysis
Posted by Haje Korth on Wed, 05 Dec 2007 16:08:02 GMT
View Forum Message <> Reply to Message

I have tried that, it gives

```
IDL> ev=imsl_princ_comp(correlate(a,/cov)) & print,ev
      45.2906      3.70938-2.65683e-006
```

These EVs are the same as you get using PCOMP with /COV keyword.

"Vince Hradil" <hradilv@yahoo.com> wrote in message
news:54fc6ed8-ccd7-4ac6-8e0d-09f5d190eeac@o6g2000hsd.googleg roups.com...
> On Dec 5, 9:12 am, Vince Hradil <hrad...@yahoo.com> wrote:
>> On Dec 5, 8:00 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
>>
>>> Hi,
>>> I am puzzled by principal component analysis. I calculated the
>>> eigenvalues
>>> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten

>>> me
>>> why the results are completely different? I have tried different
>>> keywords to
>>> see whether I can match them by trial and error, but I had no success.
>>> There
>>> must be someone out there who undertstands this much better than I do.
>>
>>> Thanks so much,
>>> Haje
>>
>>> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
>>> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
>>> 2.24227 0.757732 0.000000
>>> IDL> ev=imsl_princ_comp(a) & print,ev
>>> 9.53359 -5.19751 2.66392
>>
>> From the HELP:
>>
>> Syntax
>> Result = IMSL_PRINC_COMP(covariances [, /COV_MATRIX]
>> [, /CORR_MATRIX] [, CORRELATIONS=variable] [, CUM_PERCENT=variable] [,
>> DF=variable] [, /DOUBLE] [, EIGENVECTORS=variable] [,
>> STDEV=variable])
>>
>> Note that IMSL_PRINC_COMP requires that you pass the covariance or
>> correlation matrix - not the vectors.
>
> so maybe try
> ev=imsl_princ_comp(correlate(a,/covariance) & print, ev
> (I don't have an analyst license)

Subject: Re: Principal component analysis
Posted by Haje Korth on Wed, 05 Dec 2007 16:12:17 GMT
View Forum Message <> Reply to Message

David
I agree with that. But the Eigenvalues from the different evaluations should be somehow related. So I was looking for the relationship between the two outputs and I cannot find it.

At this point I tend to just use pcomp. As Mort showed it gives the right result and the source code is readily available. Who know what the IMSL blackbox does?!? I just got thrown off by seeing the two different results and not being able to reconcile them.

THanks,
Haje

"David Fanning" <news@dfanning.com> wrote in message
news:MPG.21c06a35c7199c9b98a129@news.frii.com...
> David Fanning writes:
>
>> Don't know. You will have to ask someone whose last math
>> class wasn't over 30 years ago. :-)
>
> I will say, though, that when I was looking into this
> I noticed that what you pass in really doesn't matter
> as long as you preserve the "vector" nature of the data.
> That is to say, the absolute magnitude of the vectors
> that comes out is irrelevant. The *relative* magnitudes
> of the orthogonal vectors "explains" the data. So I would
> say the various ways you can manipulate the data prior to
> sending it to analysis are "scaling" functions that don't
> fundamentally change the nature of the result.
>
> At least that is my hand-waving explanation. :-)
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming: http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by Vince Hradil on Wed, 05 Dec 2007 16:13:45 GMT
View Forum Message <> Reply to Message

On Dec 5, 10:08 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
> I have tried that, it gives
> IDL> ev=imsl_princ_comp(correlate(a,/cov)) & print,ev
> 45.2906 3.70938-2.65683e-006
>
> These EVs are the same as you get using PCOMP with /COV keyword.
>
> "Vince Hradil" <hrad...@yahoo.com> wrote in message
>
> news:54fc6ed8-ccd7-4ac6-8e0d-09f5d190eeac@o6g2000hsd.googleg roups.com...
>
>> On Dec 5, 9:12 am, Vince Hradil <hrad...@yahoo.com> wrote:

>>> On Dec 5, 8:00 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
>
>>>> Hi,
>>>> I am puzzled by principal component analysis. I calculated the
>>>> eigenvalues
>>>> using both PCOMP and IMSP_PRINC_COMP routines. Could someone enlighten
>>>> me
>>>> why the results are completely different? I have tried different
>>>> keywords to
>>>> see whether I can match them by trial and error, but I had no success.
>>>> There
>>>> must be someone out there who undertstands this much better than I do.
>
>>>> Thanks so much,
>>>> Haje
>
>>>> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
>>>> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
>>>> 2.24227 0.757732 0.000000
>>>> IDL> ev=imsl_princ_comp(a) & print,ev
>>>> 9.53359 -5.19751 2.66392
>
>>> From the HELP:
>
>>> Syntax
>>> Result = IMSL_PRINC_COMP(covariances [, /COV_MATRIX]
>>> [, /CORR_MATRIX] [, CORRELATIONS=variable] [, CUM_PERCENT=variable] [,
>>> DF=variable] [, /DOUBLE] [, EIGENVECTORS=variable] [,
>>> STDEV=variable])
>
>>> Note that IMSL_PRINC_COMP requires that you pass the covariance or
>>> correlation matrix - not the vectors.
>
>> so maybe try
>> ev=imsl_princ_comp(correlate(a,/covariance) & print, ev
>> (I don't have an analyst license)

There you go 8^)
How about
ev=imsl_princ_comp(correlate(a)) & print, ev

Subject: Re: Principal component analysis
Posted by Haje Korth on Wed, 05 Dec 2007 16:13:55 GMT
View Forum Message <> Reply to Message

Mort,
thanks for validating this. I tend to just go with PCOMP since I don't

really know what the IMSL routine actually does. As I wrote in the response
to David I just got thrown off by not being able to reconcile the output
from the different routines.

Thanks,
Haje

"mort canty" <m.canty@fz-juelich.de> wrote in message
news:fj6fp9$4jl8$1@zam602.zam.kfa-juelich.de...
> Haje Korth schrieb:
>> Hi,
>> I am puzzled by principal component analysis. I calculated the
>> eigenvalues using both PCOMP and IMSP_PRINC_COMP routines. Could someone
>> enlighten me why the results are completely different? I have tried
>> different keywords to see whether I can match them by trial and error,
>> but I had no success. There must be someone out there who undertstands
>> this much better than I do.
>>
>> Thanks so much,
>> Haje
>>
>>
>> IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
>> IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
>>   2.24227   0.757732   0.000000
>> IDL> ev=imsl_princ_comp(a) & print,ev
>>   9.53359   -5.19751   2.66392
>
> Haven't the foggiest what imsl_princ_comp() does, but pcomp() is correct:
>
> pro pca
> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
> ; covariance matrix
> s1 = correlate(a,/covariance)
> print, s1, ' '
> ; correlation matrix
> s2 = correlate(a)
> print, s2, ' '
> ; diagonalize
> print, eigenql(s1)
> print, eigenql(s2), ' '
> ; compare
> pca=pcomp(a,eigenvalues=ev,/covariance) & print,transpose(ev)
> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
> end
>
>   12.3333   2.33333   -19.6667
>   2.33333   4.33333   -5.66667

>   -19.6667   -5.66667   32.3333
>
>   1.00000   0.319173   -0.984839
>   0.319173   1.00000   -0.478731
>   -0.984839   -0.478731   1.00000
>
>   45.2906   3.70938-1.52795e-006
>   2.24227   0.757732-5.49480e-008
>
>   45.2906   3.70938-1.52795e-006
>   2.24227   0.757732   0.000000
>
> Mort

Subject: Re: Principal component analysis
Posted by David Fanning on Wed, 05 Dec 2007 16:25:33 GMT
View Forum Message <> Reply to Message

Haje Korth writes:

> thanks for validating this. I tend to just go with PCOMP since I don't
> really know what the IMSL routine actually does. As I wrote in the response
> to David I just got thrown off by not being able to reconcile the output
> from the different routines.

It looks to me like passing the CORRELATE results
to EIGENQL just scales the eigenvalues into -1 to 1. That
would seem to be a sensible choice to me.

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: Principal component analysis
Posted by Haje Korth on Wed, 05 Dec 2007 16:47:33 GMT
View Forum Message <> Reply to Message

Yup, that'll do it. I am still not sure I understand the logic behind this.
I though the correlation is part of the PCA.

"Vince Hradil" <hradilv@yahoo.com> wrote in message
news:8362380a-217a-45d2-b7c4-0198e5931b39@y5g2000hsf.googleg roups.com...
> On Dec 5, 10:08 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
>> I have tried that, it gives
>> IDL> ev=imsl_princ_comp(correlate(a,/cov)) & print,ev
>> 45.2906 3.70938-2.65683e-006
>>
>> These EVs are the same as you get using PCOMP with /COV keyword.
>>
>> "Vince Hradil" <hrad...@yahoo.com> wrote in message
>>
>> news:54fc6ed8-ccd7-4ac6-8e0d-09f5d190eeac@o6g2000hsd.googleg roups.com...
>>
>>> On Dec 5, 9:12 am, Vince Hradil <hrad...@yahoo.com> wrote:
>>>> On Dec 5, 8:00 am, "Haje Korth" <haje.ko...@nospam.jhuapl.edu> wrote:
>>
>>>> > Hi,
>>>> > I am puzzled by principal component analysis. I calculated the
>>>> > eigenvalues
>>>> > using both PCOMP and IMSP_PRINC_COMP routines. Could someone
>>>> > enlighten
>>>> > me
>>>> > why the results are completely different? I have tried different
>>>> > keywords to
>>>> > see whether I can match them by trial and error, but I had no
>>>> > success.
>>>> > There
>>>> > must be someone out there who undertstands this much better than I
>>>> > do.
>>
>>>> > Thanks so much,
>>>> > Haje
>>
>>>> > IDL> a=[[1,-2,-6],[-2,1,-3],[-6,-3,5]]
>>>> > IDL> pca=pcomp(a,eigenvalues=ev) & print,transpose(ev)
>>>> > 2.24227 0.757732 0.000000
>>>> > IDL> ev=imsl_princ_comp(a) & print,ev
>>>> > 9.53359 -5.19751 2.66392
>>
>>>> From the HELP:
>>
>>>> Syntax
>>>> Result = IMSL_PRINC_COMP(covariances [, /COV_MATRIX]
>>>> [, /CORR_MATRIX] [, CORRELATIONS=variable] [, CUM_PERCENT=variable] [,
>>>> DF=variable] [, /DOUBLE] [, EIGENVECTORS=variable] [,
>>>> STDEV=variable])

>>
>>>> Note that IMSL_PRINC_COMP requires that you pass the covariance or
>>>> correlation matrix - not the vectors.
>>
>>> so maybe try
>>> ev=imsl_princ_comp(correlate(a,/covariance) & print, ev
>>> (I don't have an analyst license)
>
> There you go 8^)
> How about
> ev=imsl_princ_comp(correlate(a)) & print, ev
