

PLS 回帰

青木繁伸

2019年4月27日

1 目的

5種類の方法により PLS 回帰を行う。

R の `pls` パッケージに含まれる `cppls()`, `kernelpls()`, `widekernelpls()`, `simpls()`, `oscorespls()` を Python に翻訳・修正したものである。

R の `pls` の情報

```
Package:           pls
Title:             Partial Least Squares and Principal Component Regression
Version:           2.7-0
Date:             2018-08-20
Authors@R:        c(person("Bjørn-Helge", "Mevik", role = c("aut", "cre"), email =
                  "b-h@mevik.net"), person("Ron", "Wehrens", role = "aut"), person("Kristian
                  Hovde", "Liland", role = "aut"), person("Paul", "Hiemstra", role = "ctb"))
Author:           Bjørn-Helge Mevik [aut, cre], Ron Wehrens [aut], Kristian Hovde Liland [aut],
                  Paul Hiemstra [ctb]
Maintainer:       Bjørn-Helge Mevik <b-h@mevik.net>
Encoding:         UTF-8
LazyData:         yes
Description:       Multivariate regression methods Partial Least Squares Regression (PLSR),
                  Principal Component Regression (PCR) and Canonical Powered Partial Least
                  Squares (CPPLS).
Depends:          R (>= 2.10)
Imports:          grDevices, graphics, methods, stats
Suggests:        MASS, parallel, Rmpi, testthat, RUnit
License:          GPL-2
URL:             http://mevik.net/work/software/pls.html, https://github.com/bhmevik/pls
BugReports:       https://github.com/bhmevik/pls/issues
NeedsCompilation: no
Packaged:         2018-08-20 18:36:26 UTC; bhm
Repository:       CRAN
Date/Publication: 2018-08-21 05:10:11 UTC
Built:           R 3.5.0; ; 2018-08-22 14:03:28 UTC; unix
```

参考文献

cppls	Indahl, U. (2005) A twist to partial least squares regression. <i>Journal of Chemometrics</i> , 19 , 32 – 44. Liland, K.H and Indahl, U.G (2009) Powered partial least squares discriminant analysis, <i>Journal of Chemometrics</i> , 23 , 7 – 18. Indahl, U.G., Liland, K.H. and Næs, T. (2009) Canonical partial least squares - a unified PLS approach to classification and regression problems. <i>Journal of Chemometrics</i> , 23 , 495 – 504.
kernelpls	de Jong, S. and ter Braak, C. J. F. (1994) Comments on the PLS kernel algorithm. <i>Journal of Chemometrics</i> , 8 , 169 – 174. Dayal, B. S. and MacGregor, J. F. (1997) Improved PLS algorithms. <i>Journal of Chemometrics</i> , 11 , 73 – 85.
widekernelpls	Rännar, S., Lindgren, F., Geladi, P. and Wold, S. (1994) A PLS Kernel Algorithm for Data Sets with Many Variables and Fewer Objects. Part 1: Theory and Algorithm. <i>Journal of Chemometrics</i> , 8 , 111 – 125.
simpls	de Jong, S. (1993) SIMPLS: an alternative approach to partial least squares regression. <i>Chemometrics and Intelligent Laboratory Systems</i> , 18 , 251 – 263.
oscorespls	Martens, H., Næs, T. (1989) <i>Multivariate calibration</i> . Chichester: Wiley.

2 使用法

```
import sys
sys.path.append("statlib")

from multi import cppls
cppls(x, y, ncomp, yAdd=None, center=True, lower=0.5, upper=0.5, truncPow=False,
      weights=None, verbose=True)

from multi import kernelpls
kernelpls(x, y, ncomp=None, center=True, verbose=True)

from multi import widekernelpls
widekernelpls(x, y, ncomp=None, center=True,
              tol=2.22044604925031e-16 ** 0.5, maxit=200, verbose=True)

from multi import simpls
simpls(x, y, ncomp=None, center=True, verbose=True)

from multi import oscorespls
oscorespls(x, y, ncomp=None, center=True, tol=2.22044604925031e-16 ** 0.5,
           maxit=200, verbose=True)
```

予測値, スコアのプロット

```
import sys
sys.path.append("statlib")

from multi import pls_plot
```

```
pls_plot(obj, type="p", ncomp=None, ny=1, ax1=1, ax2=2,
         txt=None, color="blue", color2="red", alpha=0.3)
```

新しいデータに対する予測値を求める

```
import sys
sys.path.append("statlib")

from multi import pls_predict
pls_predict(obj, newX=None, ncomp=None, comps=None, type="response", verbose=True):
```

2.1 引数

x	独立変数データ行列
y	従属変数データ行列 (2 列以上でもよい)
ncomp	使用する主成分の個数 (指定しなくてもよい)
yAdd	付加的な従属変数データ行列
center	デフォルトで各変数を中心化する True
lower	パワー最適化のパラメータの下限值 (デフォルトは 0.5)
upper	パワー最適化のパラメータの上限値 (デフォルトは 0.5)
truncPow	True の場合は実験的なパワーアルゴリズムを使う (デフォルトは False)
weights	各観測値の重み (デフォルトでは重み付けしない None)
verbose	必要最小限のプリント出力をする
obj	pls 回帰プログラムの戻り値
type	実測値と予測値の対比プロットの場合 (デフォルト) は "p", スコアの二次元配置をプロットする場合は "s" を指定する
ncomp	回帰に使うコンポーネント数 (デフォルトは最大数)
ny	複数の従属変数を分析した場合は何番目の従属変数かを表す (デフォルトは 1)
ax1	横軸にとるスコアの番号 (デフォルトは 1)
ax2	縦軸にとるスコアの番号 (デフォルトは 2)
color	ドットの描画色 (デフォルトは青)
txt	ドットに添えるテキスト (デフォルトは None)
color2	ドットに添えるテキストの描画色 (デフォルトは赤)
alpha	アルファチャネル (デフォルトは 0.3)
newX	新しいデータセット (独立変数データ行列のみ) デフォルト (None) の場合は分析に用いたデータセットが仮定される
ncomp	使用する合成変数の数。表示されるのはそれぞれの合成変数に対して。
comps	使用する合成変数のセット (デフォルトの場合は None)。表示されるのは総合した 1 個のモデルにおける値。
type	type="response" の場合 (デフォルト) は予測値を返す。type="score" の場合は comps で指定された合成変数の値を返す。

2.2 戻り値の名前

"coefficients"	ncomp までの主成分に対する回帰係数
"scores"	独立変数のスコア

"loadings"	独立変数の負荷量
"loadingWeights"	独立変数の負荷量の重み
"yScores"	従属変数のスコア
"yLoadings"	従属変数の負荷量
"projection"	独立変数データ行列をスコアに変換するための射影
"xMeans"	独立変数データの平均値ベクトル
"yMeans"	従属変数データの平均値ベクトル
"fittedValues"	予測値
"residuals"	残差
"xVariances"	各主成分で説明される独立変数の分散
"xTotalVariance"	独立変数の全分散 (xVariances の合計)
"gammas"	γ 値
"cancors"	正準相関係数
"A"	正準相関係数の重みベクトル
"smallNorm"	0 か 0 に近い説明変数の位置

3 使用例

3.1 3 変数を使って 1 変数を予測する例

```
import pandas as pd

df = pd.read_csv("../Python/data/iris.csv")
x = df.loc[:, ["sw", "pl", "pw"]]
y = pd.DataFrame(df.loc[:, "sl"])

import sys
sys.path.append("statlib")
from multi import cppls
a = cppls(x, y, ncomp=3)
```

```
***** Coefficients
```

```
1 comp
      sl
sw -0.011534
pl  0.346379
pw  0.140331
```

```
2 comps
      sl
sw  0.613899
pl  0.450242
pw  0.050192
```

3 comps

s1

sw 0.650837

pl 0.709132

pw -0.556483

***** Scores

	Comp1	Comp2	Comp3
Obj1	2.573133	-0.190991	-0.018997
Obj2	2.557709	0.298453	0.009857
Obj3	2.656517	0.118932	0.038761
Obj4	2.468155	0.184308	-0.036359
Obj5	2.576218	-0.288879	-0.024767
Obj6	2.232495	-0.603098	0.026142
Obj7	2.532517	-0.078994	0.081553
Obj8	2.477410	-0.109358	-0.053671
Obj9	2.554624	0.396342	0.015627
Obj10	2.505687	0.170201	-0.131138
Obj11	2.486664	-0.403024	-0.070983
Obj12	2.384771	-0.125614	-0.094117
Obj13	2.595240	0.284345	-0.084922
Obj14	2.873156	0.333113	0.036414
Obj15	2.773834	-0.647922	0.033041
Obj16	2.433196	-1.060030	0.078180
Obj17	2.603049	-0.538074	0.187924
Obj18	2.535602	-0.176883	0.075782
Obj19	2.266941	-0.519317	-0.062866
Obj20	2.452218	-0.486805	0.018025
Obj21	2.292133	-0.141870	-0.134562
Obj22	2.411602	-0.374808	0.118574
Obj23	2.946772	-0.223856	0.137014
Obj24	2.176455	-0.001657	0.155545
Obj25	2.106856	-0.174381	-0.215453
Obj26	2.372432	0.265941	-0.071034
Obj27	2.309709	-0.097398	0.095441
Obj28	2.480495	-0.207247	-0.059442
Obj29	2.570048	-0.093102	-0.013226
Obj30	2.378602	0.070164	-0.082575
...
Obj121	-2.207732	-0.300064	0.249514
Obj122	-1.366370	0.179215	0.311824
Obj123	-3.033863	-0.113392	-0.416193
Obj124	-1.294392	0.248888	0.128038
Obj125	-2.129585	-0.426169	0.054186

```

Obj126 -2.297992 -0.419371 -0.345715
Obj127 -1.198669 0.167255 0.162712
Obj128 -1.285138 -0.044779 0.110726
Obj129 -2.052371 0.079531 0.123485
Obj130 -2.043822 -0.219298 -0.442840
Obj131 -2.440501 -0.029964 -0.268299
Obj132 -2.725099 -1.043512 -0.352564
Obj133 -2.089902 0.093639 0.218264
Obj134 -1.363991 0.076164 -0.242960
Obj135 -1.795822 0.176554 -0.528424
Obj136 -2.584456 -0.169310 0.099274
Obj137 -2.146455 -0.465478 0.373197
Obj138 -1.837884 -0.240203 -0.137718
Obj139 -1.192499 -0.028523 0.151171
Obj140 -1.857839 -0.181623 0.187064
Obj141 -2.155710 -0.171812 0.390509
Obj142 -1.654986 -0.104640 0.497957
Obj143 -1.517200 0.230484 0.141925
Obj144 -2.393009 -0.332576 0.168623
Obj145 -2.279710 -0.369737 0.433301
Obj146 -1.750709 -0.023007 0.463283
Obj147 -1.430731 0.442517 0.193912
Obj148 -1.638116 -0.065331 0.178947
Obj149 -1.923647 -0.447074 0.359309
Obj150 -1.470415 -0.077290 0.029835

```

[150 rows x 3 columns]

***** Loadings

	Comp1	Comp2	Comp3
sw	0.097012	-1.029812	0.191703
pl	-0.919635	-0.007579	-0.363036
pw	-0.386534	-0.065935	0.911841

	Comp1	Comp2	Comp3
SS loadings	1.004549	1.064917	1.000000
Prop. var.	0.334850	0.354972	0.333333
Cumu. prop. var.	0.334850	0.689822	1.023156

***** Loading weithts

	Comp1	Comp2	Comp3
sw	0.030848	-0.980968	0.191703
pl	-0.926385	-0.100077	-0.363036

pw -0.375312 0.166392 0.911841

	Comp1	Comp2	Comp3
SS loadings	1.000000	1.000000	1.000000
Prop. var.	0.333333	0.333333	0.333333
Cumu. prop. var.	0.333333	0.666667	1.000000

***** y scores

	Comp1	Comp2	Comp3
Obj1	-13.639855	-9.297016	-9.260971
Obj2	-13.104959	-9.464407	-9.456908
Obj3	-12.570063	-9.029682	-9.051884
Obj4	-12.302615	-8.828314	-8.775766
Obj5	-13.372407	-9.044419	-9.003067
Obj6	-14.442200	-9.155104	-9.164458
Obj7	-12.302615	-8.602677	-8.668454
Obj8	-13.372407	-9.166117	-9.095637
Obj9	-11.767718	-8.777923	-8.777454
Obj10	-13.104959	-9.305710	-9.157508
Obj11	-14.442200	-9.503921	-9.415509
Obj12	-12.837511	-8.782622	-8.672400
Obj13	-12.837511	-9.315749	-9.213744
Obj14	-11.500270	-8.744589	-8.764968
Obj15	-15.511992	-10.053131	-10.067736
Obj16	-15.244544	-9.285165	-9.346317
Obj17	-14.442200	-9.436980	-9.607598
Obj18	-13.639855	-9.289161	-9.347908
Obj19	-15.244544	-9.728584	-9.647877
Obj20	-13.639855	-8.930441	-8.932089
Obj21	-14.442200	-9.651233	-9.498973
Obj22	-13.639855	-9.018669	-9.120705
Obj23	-12.302615	-8.700242	-8.821302
Obj24	-13.639855	-9.254209	-9.392783
Obj25	-12.837511	-8.571215	-8.340044
Obj26	-13.372407	-9.479982	-9.391564
Obj27	-13.372407	-9.079937	-9.158727
Obj28	-13.907303	-9.383061	-9.306412
Obj29	-13.907303	-9.549613	-9.518875
Obj30	-12.570063	-8.818275	-8.719529
...
Obj121	-18.453922	-9.207372	-9.440002
Obj122	-14.977096	-8.144351	-8.441240
Obj123	-20.593507	-10.162692	-9.727862
Obj124	-16.849233	-9.351740	-9.462628

```

Obj125 -17.919026 -8.813974 -8.851997
Obj126 -19.256266 -9.504785 -9.141640
Obj127 -16.581785 -9.169612 -9.315509
Obj128 -16.314337 -8.750463 -8.845142
Obj129 -17.116681 -8.895320 -9.002493
Obj130 -19.256266 -9.853601 -9.392691
Obj131 -19.791163 -10.123821 -9.836956
Obj132 -21.128403 -9.726290 -9.355891
Obj133 -17.116681 -8.887464 -9.089430
Obj134 -16.849233 -9.138286 -8.878568
Obj135 -16.314337 -8.672937 -8.128609
Obj136 -20.593507 -10.369771 -10.450029
Obj137 -16.849233 -8.138738 -8.497011
Obj138 -17.116681 -8.701106 -8.547432
Obj139 -16.046889 -8.664418 -8.799700
Obj140 -18.453922 -9.530574 -9.700161
Obj141 -17.919026 -9.053042 -9.426950
Obj142 -18.453922 -9.726269 -10.206391
Obj143 -15.511992 -8.420379 -8.546863
Obj144 -18.186474 -8.909921 -9.062206
Obj145 -17.919026 -8.782550 -9.199746
Obj146 -17.919026 -9.438857 -9.884831
Obj147 -16.849233 -9.465582 -9.642136
Obj148 -17.384129 -9.149398 -9.311566
Obj149 -16.581785 -8.131018 -8.475417
Obj150 -15.779440 -8.296498 -8.311119

```

[150 rows x 3 columns]

***** y loadings

	Comp1	Comp2	Comp3
sl	-0.373904	-0.638923	-0.640097
	Comp1	Comp2	Comp3
SS loadings	0.139804	0.408223	0.409724
Prop. var.	0.139804	0.408223	0.409724
Cumu. prop. var.	0.139804	0.548027	0.957751

Projection

	Comp1	Comp2	Comp3
sw	0.030848	-0.978887	-0.057707
pl	-0.926385	-0.162559	-0.404454
pw	-0.375312	0.141078	0.947787

***** Fitted values

1 comp	s1
Obj1	4.881228
Obj2	4.886995
Obj3	4.850050
Obj4	4.920479
Obj5	4.880074
Obj6	5.008594
Obj7	4.896414
Obj8	4.917019
Obj9	4.888148
Obj10	4.906446
Obj11	4.913559
Obj12	4.951657
Obj13	4.872962
Obj14	4.769048
Obj15	4.806185
Obj16	4.933551
Obj17	4.870042
Obj18	4.895261
Obj19	4.995714
Obj20	4.926439
Obj21	4.986295
Obj22	4.941625
Obj23	4.741523
Obj24	5.029548
Obj25	5.055571
Obj26	4.956271
Obj27	4.979723
Obj28	4.915866
Obj29	4.882381
Obj30	4.953964
...	...
Obj121	6.668814
Obj122	6.354225
Obj123	6.977708
Obj124	6.327312
Obj125	6.639594
Obj126	6.702562
Obj127	6.291521
Obj128	6.323852
Obj129	6.610723

Obj130 6.607527
Obj131 6.755847
Obj132 6.862260
Obj133 6.624756
Obj134 6.353335
Obj135 6.514799
Obj136 6.809672
Obj137 6.645902
Obj138 6.530526
Obj139 6.289214
Obj140 6.537987
Obj141 6.649362
Obj142 6.462140
Obj143 6.410621
Obj144 6.738090
Obj145 6.695726
Obj146 6.497931
Obj147 6.378290
Obj148 6.455832
Obj149 6.562593
Obj150 6.393128

[150 rows x 1 columns]

2 comps

 s1
Obj1 5.003256
Obj2 4.696306
Obj3 4.774062
Obj4 4.802721
Obj5 5.064646
Obj6 5.393927
Obj7 4.946885
Obj8 4.986890
Obj9 4.634917
Obj10 4.797701
Obj11 5.171060
Obj12 5.031915
Obj13 4.691287
Obj14 4.556215
Obj15 5.220157
Obj16 5.610828
Obj17 5.213830
Obj18 5.008275
Obj19 5.327518

Obj20	5.237469
Obj21	5.076939
Obj22	5.181099
Obj23	4.884549
Obj24	5.030607
Obj25	5.166987
Obj26	4.786355
Obj27	5.041953
Obj28	5.048280
Obj29	4.941866
Obj30	4.909135
...	...
Obj121	6.860532
Obj122	6.239721
Obj123	7.050156
Obj124	6.168292
Obj125	6.911883
Obj126	6.970508
Obj127	6.184658
Obj128	6.352462
Obj129	6.559909
Obj130	6.747641
Obj131	6.774992
Obj132	7.528983
Obj133	6.564929
Obj134	6.304673
Obj135	6.401995
Obj136	6.917849
Obj137	6.943307
Obj138	6.683997
Obj139	6.307438
Obj140	6.654031
Obj141	6.759137
Obj142	6.528997
Obj143	6.263360
Obj144	6.950580
Obj145	6.931960
Obj146	6.512631
Obj147	6.095556
Obj148	6.497573
Obj149	6.848239
Obj150	6.442510

[150 rows x 1 columns]

3 comps

	s1
Obj1	5.015416
Obj2	4.689997
Obj3	4.749251
Obj4	4.825994
Obj5	5.080499
Obj6	5.377194
Obj7	4.894684
Obj8	5.021245
Obj9	4.624913
Obj10	4.881642
Obj11	5.216496
Obj12	5.092158
Obj13	4.745645
Obj14	4.532906
Obj15	5.199008
Obj16	5.560786
Obj17	5.093541
Obj18	4.959767
Obj19	5.367758
Obj20	5.225932
Obj21	5.163072
Obj22	5.105200
Obj23	4.796847
Obj24	4.931043
Obj25	5.304898
Obj26	4.831824
Obj27	4.980862
Obj28	5.086329
Obj29	4.950332
Obj30	4.961991
...	...
Obj121	6.700818
Obj122	6.040123
Obj123	7.316560
Obj124	6.086336
Obj125	6.877199
Obj126	7.191799
Obj127	6.080506
Obj128	6.281587
Obj129	6.480867
Obj130	7.031102
Obj131	6.946729
Obj132	7.754658

Obj133 6.425219
Obj134 6.460191
Obj135 6.740237
Obj136 6.854304
Obj137 6.704424
Obj138 6.772150
Obj139 6.210674
Obj140 6.534292
Obj141 6.509173
Obj142 6.210256
Obj143 6.172514
Obj144 6.842645
Obj145 6.654606
Obj146 6.216085
Obj147 5.971433
Obj148 6.383030
Obj149 6.618246
Obj150 6.423413

[150 rows x 1 columns]

***** Residuals

1 comp

	s1
Obj1	0.218772
Obj2	0.013005
Obj3	-0.150050
Obj4	-0.320479
Obj5	0.119926
Obj6	0.391406
Obj7	-0.296414
Obj8	0.082981
Obj9	-0.488148
Obj10	-0.006446
Obj11	0.486441
Obj12	-0.151657
Obj13	-0.072962
Obj14	-0.469048
Obj15	0.993815
Obj16	0.766449
Obj17	0.529958
Obj18	0.204739
Obj19	0.704286
Obj20	0.173561

Obj21 0.413705
Obj22 0.158375
Obj23 -0.141523
Obj24 0.070452
Obj25 -0.255571
Obj26 0.043729
Obj27 0.020277
Obj28 0.284134
Obj29 0.317619
Obj30 -0.253964
... ..
Obj121 0.231186
Obj122 -0.754225
Obj123 0.722292
Obj124 -0.027312
Obj125 0.060406
Obj126 0.497438
Obj127 -0.091521
Obj128 -0.223852
Obj129 -0.210723
Obj130 0.592473
Obj131 0.644153
Obj132 1.037740
Obj133 -0.224756
Obj134 -0.053335
Obj135 -0.414799
Obj136 0.890328
Obj137 -0.345902
Obj138 -0.130526
Obj139 -0.289214
Obj140 0.362013
Obj141 0.050638
Obj142 0.437860
Obj143 -0.610621
Obj144 0.061910
Obj145 0.004274
Obj146 0.202069
Obj147 -0.078290
Obj148 0.044168
Obj149 -0.362593
Obj150 -0.493128

[150 rows x 1 columns]

2 comps

s1
Obj1 0.096744
Obj2 0.203694
Obj3 -0.074062
Obj4 -0.202721
Obj5 -0.064646
Obj6 0.006073
Obj7 -0.346885
Obj8 0.013110
Obj9 -0.234917
Obj10 0.102299
Obj11 0.228940
Obj12 -0.231915
Obj13 0.108713
Obj14 -0.256215
Obj15 0.579843
Obj16 0.089172
Obj17 0.186170
Obj18 0.091725
Obj19 0.372482
Obj20 -0.137469
Obj21 0.323061
Obj22 -0.081099
Obj23 -0.284549
Obj24 0.069393
Obj25 -0.366987
Obj26 0.213645
Obj27 -0.041953
Obj28 0.151720
Obj29 0.258134
Obj30 -0.209135
... ..
Obj121 0.039468
Obj122 -0.639721
Obj123 0.649844
Obj124 0.131708
Obj125 -0.211883
Obj126 0.229492
Obj127 0.015342
Obj128 -0.252462
Obj129 -0.159909
Obj130 0.452359
Obj131 0.625008
Obj132 0.371017
Obj133 -0.164929

Obj134 -0.004673
Obj135 -0.301995
Obj136 0.782151
Obj137 -0.643307
Obj138 -0.283997
Obj139 -0.307438
Obj140 0.245969
Obj141 -0.059137
Obj142 0.371003
Obj143 -0.463360
Obj144 -0.150580
Obj145 -0.231960
Obj146 0.187369
Obj147 0.204444
Obj148 0.002427
Obj149 -0.648239
Obj150 -0.542510

[150 rows x 1 columns]

3 comps

 s1
Obj1 0.084584
Obj2 0.210003
Obj3 -0.049251
Obj4 -0.225994
Obj5 -0.080499
Obj6 0.022806
Obj7 -0.294684
Obj8 -0.021245
Obj9 -0.224913
Obj10 0.018358
Obj11 0.183504
Obj12 -0.292158
Obj13 0.054355
Obj14 -0.232906
Obj15 0.600992
Obj16 0.139214
Obj17 0.306459
Obj18 0.140233
Obj19 0.332242
Obj20 -0.125932
Obj21 0.236928
Obj22 -0.005200
Obj23 -0.196847


```
Obj24  0.168957
Obj25 -0.504898
Obj26  0.168176
Obj27  0.019138
Obj28  0.113671
Obj29  0.249668
Obj30 -0.261991
...
Obj121 0.199182
Obj122 -0.440123
Obj123 0.383440
Obj124 0.213664
Obj125 -0.177199
Obj126 0.008201
Obj127 0.119494
Obj128 -0.181587
Obj129 -0.080867
Obj130 0.168898
Obj131 0.453271
Obj132 0.145342
Obj133 -0.025219
Obj134 -0.160191
Obj135 -0.640237
Obj136 0.845696
Obj137 -0.404424
Obj138 -0.372150
Obj139 -0.210674
Obj140 0.365708
Obj141 0.190827
Obj142 0.689744
Obj143 -0.372514
Obj144 -0.042645
Obj145 0.045394
Obj146 0.483915
Obj147 0.328567
Obj148 0.116970
Obj149 -0.418246
Obj150 -0.523413
```

```
[150 rows x 1 columns]
```

```
***** x means
```

```
sw  3.057333
```

```
pl  3.758000
```

```
pw  1.199333
```

```
***** y means
sl 5.843333

***** Explained variances of x by each component
Comp1 550.658541
Comp2 23.041092
Comp3 5.502634

***** Total variance of x
579.20227

***** Gammas
[0.5 0.5 0.5]

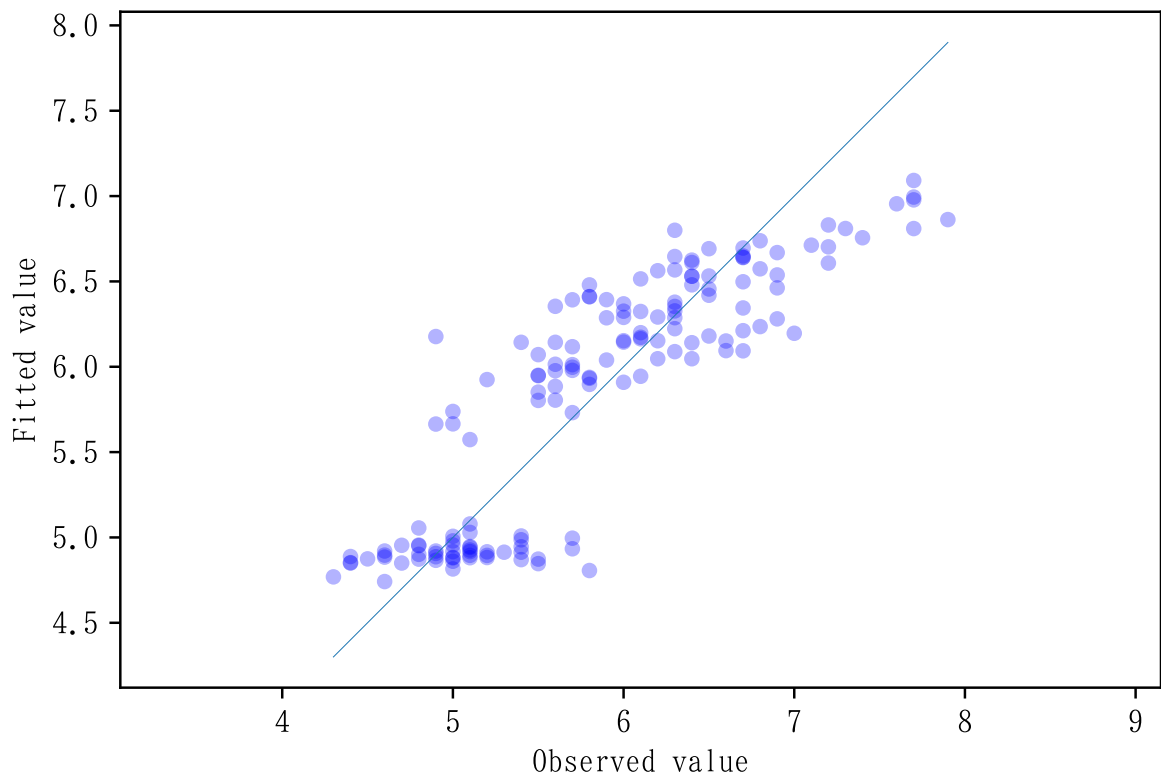
***** Canonical correlations
[0.01467038 0.0016908 0.00043159]

***** A
[[-0.0025437 -0.18982961 -1.47737782]]
```

実測値と予測値の関係図
合成変数1個を使う場合

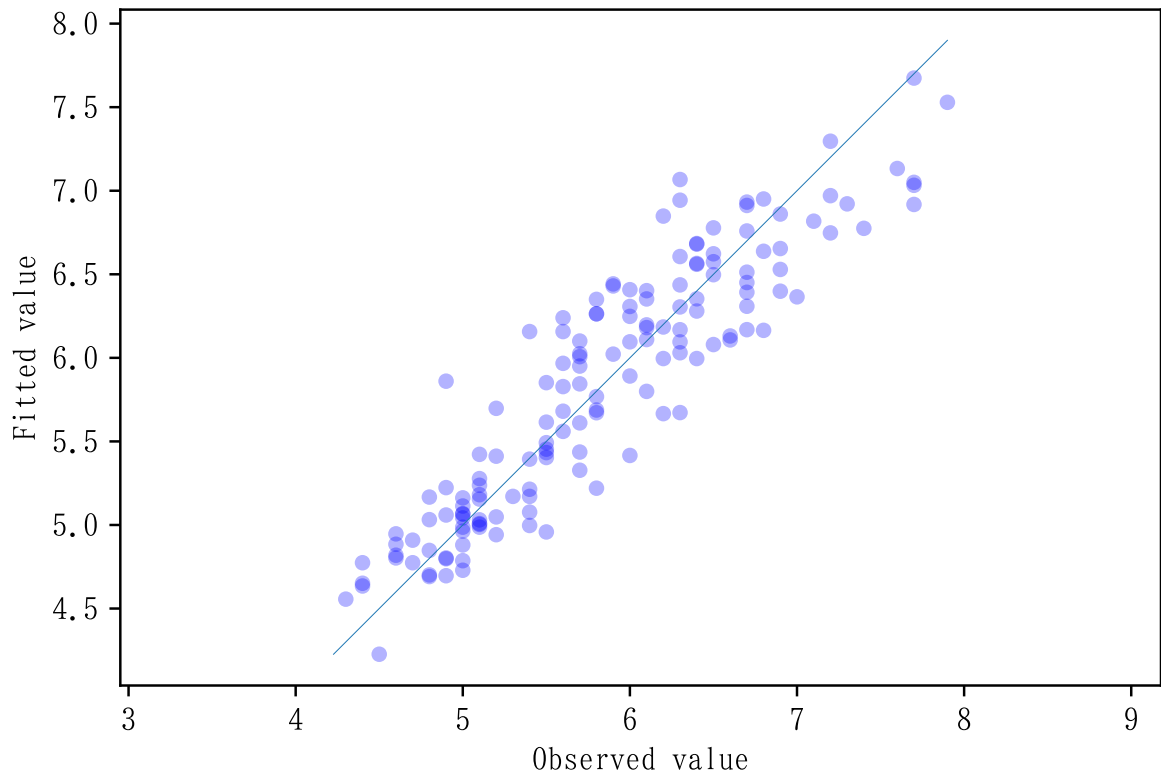
```
import sys
sys.path.append("statlib")
from multi import pls_plot

pls_plot(a, ncomp=1)
```



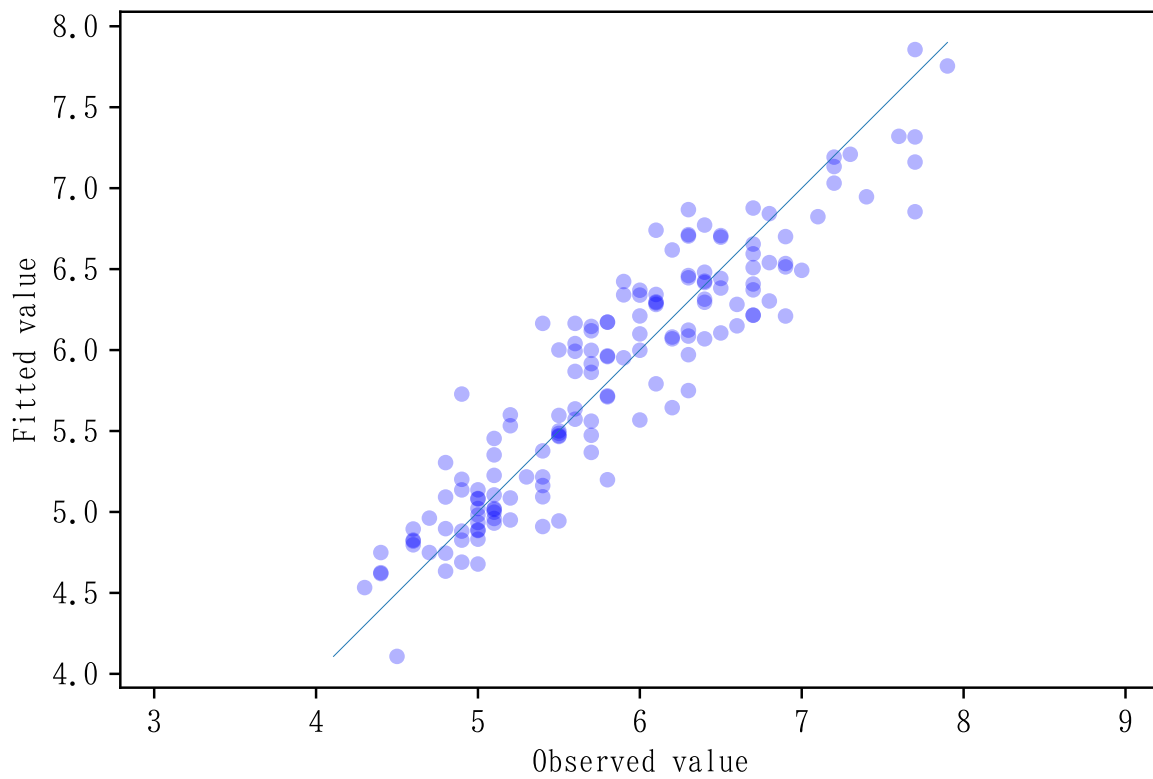
合成変数 2 個を使う場合

```
pls_plot(a, ncomp=2)
```

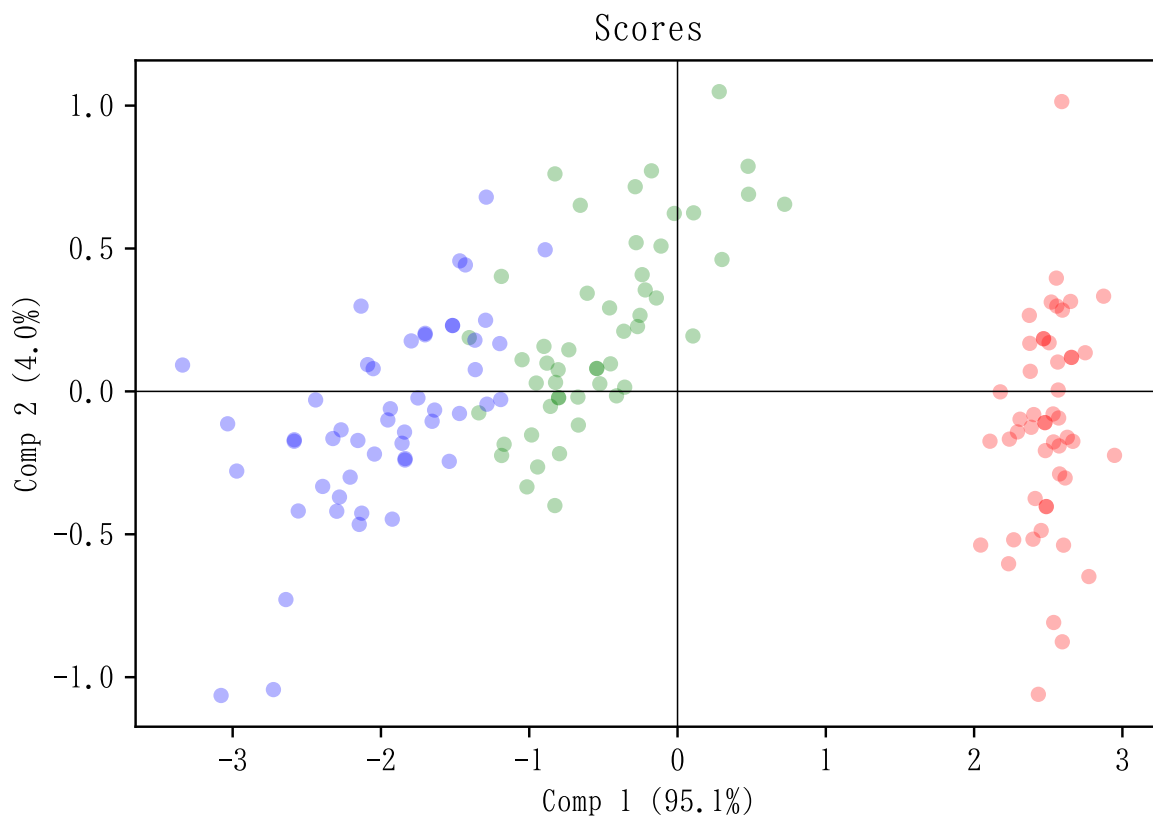


合成変数 3 個を使う場合

```
pls_plot(a, ncomp=3)
```



```
import scipy as sp
color = sp.hstack((sp.repeat("red", 50), sp.repeat("green", 50), sp.
    repeat("blue", 50)))
pls_plot(a, type="s", ncomp=[1, 2], color=color)
```



3.2 2変数を使って2変数を予測する例

```
df = pd.read_csv("../Python/data/iris.csv")
x = df.iloc[:, :2]
y = df.iloc[:, 2:4]
a = cppls(x, y, ncomp=2)
```

***** Coefficients

1 comp

	p1	pw
sl	1.863112	0.747580
sw	-0.866620	-0.347734

2 comps

	p1	pw
sl	1.775593	0.723292
sw	-1.338623	-0.478721

***** Scores

	Comp1	Comp2
Obj1	0.860684	-0.309506
Obj2	0.831150	0.235805
Obj3	1.096843	0.070394
Obj4	1.145338	0.190752
Obj5	0.993531	-0.392212
Obj6	0.757372	-0.772112
Obj7	1.271864	-0.113843
Obj8	0.909180	-0.189148
Obj9	1.242330	0.431468
Obj10	0.873325	0.134274
Obj11	0.673022	-0.569048
Obj12	1.090522	-0.151496
Obj13	0.921821	0.254632
Obj14	1.375176	0.348763
Obj15	0.436864	-0.948949
Obj16	0.696236	-1.336250
Obj17	0.757372	-0.772112
Obj18	0.860684	-0.309506
Obj19	0.443184	-0.727059
Obj20	0.987210	-0.614102
Obj21	0.546496	-0.264453
Obj22	0.945035	-0.512570

Obj23	1.356215	-0.316907
Obj24	0.776334	-0.106443
Obj25	1.090522	-0.151496
Obj26	0.740479	0.216979
Obj27	0.909180	-0.189148
Obj28	0.770013	-0.328332
Obj29	0.727838	-0.226801
Obj30	1.096843	0.070394
...
Obj121	-0.897921	-0.343782
Obj122	0.112102	0.307086
Obj123	-1.791991	-0.088264
Obj124	-0.564771	0.276834
Obj125	-0.674403	-0.407662
Obj126	-1.169934	-0.400261
Obj127	-0.431925	0.194128
Obj128	-0.256903	0.009891
Obj129	-0.613267	0.156476
Obj130	-1.254285	-0.197197
Obj131	-1.519977	-0.031786
Obj132	-1.551579	-1.141235
Obj133	-0.613267	0.156476
Obj134	-0.522596	0.175302
Obj135	-0.425604	0.416018
Obj136	-1.707640	-0.291328
Obj137	-0.269544	-0.433889
Obj138	-0.486741	-0.148119
Obj139	-0.166232	0.028717
Obj140	-0.940096	-0.242250
Obj141	-0.758754	-0.204598
Obj142	-0.940096	-0.242250
Obj143	-0.111416	0.370965
Obj144	-0.807250	-0.324956
Obj145	-0.674403	-0.407662
Obj146	-0.800929	-0.103066
Obj147	-0.649122	0.479898
Obj148	-0.619587	-0.065414
Obj149	-0.178873	-0.415063
Obj150	-0.075561	0.047543

[150 rows x 2 columns]

***** Loadings

	Comp1	Comp2
--	-------	-------

```
sl -1.015318 -0.421753
sw 0.188262 -0.906711
```

	Comp1	Comp2
SS loadings	1.066314	1.000000
Prop. var.	0.533157	0.500000
Cumu. prop. var.	0.533157	1.033157

***** Loading weithts

	Comp1	Comp2
sl	-0.906711	-0.421753
sw	0.421753	-0.906711

	Comp1	Comp2
SS loadings	1.0	1.0
Prop. var.	0.5	0.5
Cumu. prop. var.	0.5	1.0

***** y scores

	Comp1	Comp2
Obj1	-0.620485	7.142100
Obj2	-0.620485	6.462082
Obj3	-0.578567	7.639587
Obj4	-0.662402	8.139871
Obj5	-0.620485	7.830716
Obj6	-0.779877	7.843536
Obj7	-0.637304	8.877249
Obj8	-0.662402	7.442657
Obj9	-0.620485	8.141804
Obj10	-0.645583	6.900275
Obj11	-0.662402	6.745443
Obj12	-0.704320	8.431830
Obj13	-0.603665	6.801378
Obj14	-0.477913	8.175813
Obj15	-0.536650	5.449050
Obj16	-0.696042	7.729378
Obj17	-0.612206	7.044629
Obj18	-0.637304	7.197527
Obj19	-0.763057	6.310047
Obj20	-0.679222	8.278932
Obj21	-0.746238	6.263218
Obj22	-0.696042	8.040466
Obj23	-0.452815	8.610701

Obj24	-0.796696	7.319775
Obj25	-0.830073	9.031010
Obj26	-0.704320	6.466812
Obj27	-0.737959	7.753237
Obj28	-0.662402	6.947104
Obj29	-0.620485	6.453484
Obj30	-0.704320	8.238767
...
Obj121	-2.776150	8.907619
Obj122	-2.390351	11.099352
Obj123	-3.144867	6.405250
Obj124	-2.356712	7.931545
Obj125	-2.742511	9.880104
Obj126	-2.817805	8.045496
Obj127	-2.314795	8.420434
Obj128	-2.356712	9.602669
Obj129	-2.700594	9.395082
Obj130	-2.700331	6.947403
Obj131	-2.876542	6.335632
Obj132	-3.019115	7.955552
Obj133	-2.717413	9.450509
Obj134	-2.390089	8.458611
Obj135	-2.582857	9.603478
Obj136	-2.943821	5.960956
Obj137	-2.751052	11.719442
Obj138	-2.608217	9.910753
Obj139	-2.314795	9.797666
Obj140	-2.616759	7.903692
Obj141	-2.751052	9.258872
Obj142	-2.524645	7.415366
Obj143	-2.457367	10.360040
Obj144	-2.859985	9.701795
Obj145	-2.809789	10.101811
Obj146	-2.566563	8.110645
Obj147	-2.415449	7.598913
Obj148	-2.516104	8.733811
Obj149	-2.650398	11.659285
Obj150	-2.440547	10.791569

[150 rows x 2 columns]

***** y loadings

	Comp1	Comp2
p1	-2.054803	0.464882

pw -0.824497 0.129011

	Comp1	Comp2
SS loadings	4.902012	0.232759
Prop. var.	2.451006	0.116380
Cumu. prop. var.	2.451006	2.567386

Projection

	Comp1	Comp2
sl	-0.906711	-0.188262
sw	0.421753	-1.015318

***** Fitted values

1 comp

	pl	pw
Obj1	1.989463	0.489702
Obj2	2.050150	0.514053
Obj3	1.504204	0.294990
Obj4	1.404555	0.255006
Obj5	1.716490	0.380171
Obj6	2.201748	0.574882
Obj7	1.144569	0.150685
Obj8	1.889814	0.449717
Obj9	1.205256	0.175037
Obj10	1.963488	0.479280
Obj11	2.375072	0.644429
Obj12	1.517191	0.300201
Obj13	1.863839	0.439295
Obj14	0.932283	0.065505
Obj15	2.860331	0.839141
Obj16	2.327372	0.625289
Obj17	2.201748	0.574882
Obj18	1.989463	0.489702
Obj19	2.847344	0.833930
Obj20	1.729477	0.385382
Obj21	2.635058	0.748749
Obj22	1.816139	0.420155
Obj23	0.971245	0.081139
Obj24	2.162787	0.559249
Obj25	1.517191	0.300201
Obj26	2.236462	0.588811
Obj27	1.889814	0.449717
Obj28	2.175774	0.564460

Obj29	2.262436	0.599233
Obj30	1.504204	0.294990
...
Obj121	5.603051	1.939666
Obj122	3.527653	1.106906
Obj123	7.440188	2.676824
Obj124	4.918493	1.664985
Obj125	5.143766	1.755377
Obj126	6.161984	2.163940
Obj127	4.645520	1.555454
Obj128	4.285885	1.411149
Obj129	5.018143	1.704970
Obj130	6.335308	2.233487
Obj131	6.881255	2.452550
Obj132	6.946191	2.478605
Obj133	5.018143	1.704970
Obj134	4.831831	1.630212
Obj135	4.632533	1.550243
Obj136	7.266864	2.607277
Obj137	4.311859	1.421571
Obj138	4.758157	1.600649
Obj139	4.099574	1.336391
Obj140	5.689713	1.974439
Obj141	5.317090	1.824923
Obj142	5.689713	1.974439
Obj143	3.986937	1.291195
Obj144	5.416739	1.864908
Obj145	5.143766	1.755377
Obj146	5.403752	1.859697
Obj147	5.091817	1.734532
Obj148	5.031130	1.710181
Obj149	4.125548	1.346813
Obj150	3.913263	1.261633

[150 rows x 2 columns]

2 comps

	pl	pw
Obj1	1.845579	0.449772
Obj2	2.159772	0.544475
Obj3	1.536929	0.304072
Obj4	1.493232	0.279615
Obj5	1.534157	0.329571
Obj6	1.842807	0.475271
Obj7	1.091645	0.135998

Obj8	1.801882	0.425315
Obj9	1.405838	0.230701
Obj10	2.025910	0.496602
Obj11	2.110532	0.571016
Obj12	1.446764	0.280657
Obj13	1.982213	0.472145
Obj14	1.094417	0.110499
Obj15	2.419182	0.716716
Obj16	1.706174	0.452898
Obj17	1.842807	0.475271
Obj18	1.845579	0.449772
Obj19	2.509348	0.740131
Obj20	1.443992	0.306156
Obj21	2.512119	0.714632
Obj22	1.577854	0.354028
Obj23	0.823920	0.040254
Obj24	2.113304	0.545517
Obj25	1.446764	0.280657
Obj26	2.337331	0.616804
Obj27	1.801882	0.425315
Obj28	2.023138	0.522102
Obj29	2.157001	0.569974
Obj30	1.536929	0.304072
...
Obj121	5.443233	1.895314
Obj122	3.670412	1.146523
Obj123	7.399156	2.665436
Obj124	5.047189	1.700700
Obj125	4.954252	1.702784
Obj126	5.975910	2.112302
Obj127	4.735767	1.580498
Obj128	4.290483	1.412425
Obj129	5.090886	1.725157
Obj130	6.243635	2.208046
Obj131	6.866478	2.448449
Obj132	6.415651	2.331374
Obj133	5.090886	1.725157
Obj134	4.913326	1.652828
Obj135	4.825932	1.603913
Obj136	7.131431	2.569692
Obj137	4.110152	1.365595
Obj138	4.689299	1.581540
Obj139	4.112924	1.340096
Obj140	5.577095	1.943186
Obj141	5.221976	1.798528

Obj142	5.577095	1.943186
Obj143	4.159392	1.339054
Obj144	5.265673	1.822985
Obj145	4.954252	1.702784
Obj146	5.355839	1.846400
Obj147	5.314913	1.796444
Obj148	5.000720	1.701742
Obj149	3.932593	1.293266
Obj150	3.935365	1.267767

[150 rows x 2 columns]

***** Residuals

1 comp

	pl	pw
Obj1	-0.589463	-0.289702
Obj2	-0.650150	-0.314053
Obj3	-0.204204	-0.094990
Obj4	0.095445	-0.055006
Obj5	-0.316490	-0.180171
Obj6	-0.501748	-0.174882
Obj7	0.255431	0.149315
Obj8	-0.389814	-0.249717
Obj9	0.194744	0.024963
Obj10	-0.463488	-0.379280
Obj11	-0.875072	-0.444429
Obj12	0.082809	-0.100201
Obj13	-0.463839	-0.339295
Obj14	0.167717	0.034495
Obj15	-1.660331	-0.639141
Obj16	-0.827372	-0.225289
Obj17	-0.901748	-0.174882
Obj18	-0.589463	-0.189702
Obj19	-1.147344	-0.533930
Obj20	-0.229477	-0.085382
Obj21	-0.935058	-0.548749
Obj22	-0.316139	-0.020155
Obj23	0.028755	0.118861
Obj24	-0.462787	-0.059249
Obj25	0.382809	-0.100201
Obj26	-0.636462	-0.388811
Obj27	-0.289814	-0.049717
Obj28	-0.675774	-0.364460
Obj29	-0.862436	-0.399233

Obj30	0.095796	-0.094990
...
Obj121	0.096949	0.360334
Obj122	1.372347	0.893094
Obj123	-0.740188	-0.676824
Obj124	-0.018493	0.135015
Obj125	0.556234	0.344623
Obj126	-0.161984	-0.363940
Obj127	0.154480	0.244546
Obj128	0.614115	0.388851
Obj129	0.581857	0.395030
Obj130	-0.535308	-0.633487
Obj131	-0.781255	-0.552550
Obj132	-0.546191	-0.478605
Obj133	0.581857	0.495030
Obj134	0.268169	-0.130212
Obj135	0.967467	-0.150243
Obj136	-1.166864	-0.307277
Obj137	1.288141	0.978429
Obj138	0.741843	0.199351
Obj139	0.700426	0.463609
Obj140	-0.289713	0.125561
Obj141	0.282910	0.575077
Obj142	-0.589713	0.325561
Obj143	1.113063	0.608805
Obj144	0.483261	0.435092
Obj145	0.556234	0.744623
Obj146	-0.203752	0.440303
Obj147	-0.091817	0.165468
Obj148	0.168870	0.289819
Obj149	1.274452	0.953187
Obj150	1.186737	0.538367

[150 rows x 2 columns]

2 comps

	pl	pw
Obj1	-0.445579	-0.249772
Obj2	-0.759772	-0.344475
Obj3	-0.236929	-0.104072
Obj4	0.006768	-0.079615
Obj5	-0.134157	-0.129571
Obj6	-0.142807	-0.075271
Obj7	0.308355	0.164002
Obj8	-0.301882	-0.225315

Obj9 -0.005838 -0.030701
Obj10 -0.525910 -0.396602
Obj11 -0.610532 -0.371016
Obj12 0.153236 -0.080657
Obj13 -0.582213 -0.372145
Obj14 0.005583 -0.010499
Obj15 -1.219182 -0.516716
Obj16 -0.206174 -0.052898
Obj17 -0.542807 -0.075271
Obj18 -0.445579 -0.149772
Obj19 -0.809348 -0.440131
Obj20 0.056008 -0.006156
Obj21 -0.812119 -0.514632
Obj22 -0.077854 0.045972
Obj23 0.176080 0.159746
Obj24 -0.413304 -0.045517
Obj25 0.453236 -0.080657
Obj26 -0.737331 -0.416804
Obj27 -0.201882 -0.025315
Obj28 -0.523138 -0.322102
Obj29 -0.757001 -0.369974
Obj30 0.063071 -0.104072
... ..
Obj121 0.256767 0.404686
Obj122 1.229588 0.853477
Obj123 -0.699156 -0.665436
Obj124 -0.147189 0.099300
Obj125 0.745748 0.397216
Obj126 0.024090 -0.312302
Obj127 0.064233 0.219502
Obj128 0.609517 0.387575
Obj129 0.509114 0.374843
Obj130 -0.443635 -0.608046
Obj131 -0.766478 -0.548449
Obj132 -0.015651 -0.331374
Obj133 0.509114 0.474843
Obj134 0.186674 -0.152828
Obj135 0.774068 -0.203913
Obj136 -1.031431 -0.269692
Obj137 1.489848 1.034405
Obj138 0.810701 0.218460
Obj139 0.687076 0.459904
Obj140 -0.177095 0.156814
Obj141 0.378024 0.601472
Obj142 -0.477095 0.356814

```
Obj143 0.940608 0.560946
Obj144 0.634327 0.477015
Obj145 0.745748 0.797216
Obj146 -0.155839 0.453600
Obj147 -0.314913 0.103556
Obj148 0.199280 0.298258
Obj149 1.467407 1.006734
Obj150 1.164635 0.532233
```

```
[150 rows x 2 columns]
```

```
***** x means
```

```
sl 5.843333
```

```
sw 3.057333
```

```
***** y means
```

```
pl 3.758000
```

```
pw 1.199333
```

```
***** Explained variances of x by each component
```

```
Comp1 100.090457
```

```
Comp2 30.384809
```

```
***** Total variance of x
```

```
130.47527
```

```
***** Gammas
```

```
[0.5 0.5]
```

```
***** Canonical correlations
```

```
[0.32184625 0.00485945]
```

```
***** A
```

```
[[-0.05983197 0.15677097]
```

```
[ 0.13283316 0.          ]]
```

3.3 新しいデータセットに対する予測

cpppls() による分析結果を obj に代入する。

```
import scipy as sp
import pandas as pd

df = pd.read_csv("../Python/data/pls.csv")
x = df.loc[:, ["X1", "X2", "X3", "X4", "X5"]]
y = df.loc[:, ["Y1", "Y2", "Y3"]]
```

```
import sys
sys.path.append("statlib")
from multi import cppls, pls_predict
obj = cppls(x, y, verbose=False)
```

新しいデータセット ($n = 4$)

```
newX = [[71.5, 39.5, 71.6, 46.1, 41.9],
        [38.4, 40.9, 47.2, 50.9, 53.9],
        [49.8, 54.2, 41.7, 27.0, 72.8],
        [53.0, 42.3, 71.9, 64.7, 65.2]]
```

`ncomp` を省略すると、理論的に可能な合成変数の数までのそれぞれについて予測値を求める。

この場合は合成変数を 1 個だけ、2 個まで、..., 5 個までに対して予測値を求める。1 ~ 3 列は、Y1, Y2, Y3 の 3 変数に対する予測値である。行数 4 は $n = 4$ に対応する。4 行列が `comp=5` 個表示される。

```
a = pls_predict(obj, newX=newX)
```

***** Predicted values

```
[[[53.93437133 52.42366065 73.25389697]
  [48.84498667 49.28798941 36.71220461]
  [46.32095657 47.7328842 18.58971283]
  [50.56297354 50.34647533 49.04732009]]
```

```
[[[58.33638332 51.93717508 74.24251925]
  [49.63206424 49.20100603 36.88896979]
  [52.36319954 47.06512954 19.94670506]
  [55.40388662 49.81148489 50.13451262]]
```

```
[[[57.42496753 51.11311248 76.46305091]
  [49.47792779 49.06164254 37.26450078]
  [52.35734184 47.05983326 19.96097649]
  [55.81056829 50.17918881 49.14369202]]
```

```
[[[60.16128321 53.12409669 78.652381 ]
  [46.00883968 46.51212623 34.48887842]
  [55.84530954 49.62322463 22.75170444]
  [59.0248579 52.54144757 51.71544973]]
```

```
[[[59.81989093 52.70381925 78.30231213]
  [47.55825186 48.41955949 36.07766941]
  [53.11102888 46.25713657 19.94793116]
  [64.74794061 59.58695707 57.58398641]]]
```

4 個までの合成変数を使った場合 (`ncomp=4`) の予測値を求める。

```
a = pls_predict(obj, newX=newX, ncomp=4)
```



```
***** Predicted values
[[[60.16128321 53.12409669 78.652381 ]
  [46.00883968 46.51212623 34.48887842]
  [55.84530954 49.62322463 22.75170444]
  [59.0248579 52.54144757 51.71544973]]]
```

1 個だけの合成変数を使った場合、2 個までの合成変数を使った場合 (ncomp=4) の予測値を求める。

```
a = pls_predict(obj, newX=newX, ncomp=[1,2])
```

```
***** Predicted values
[[[53.93437133 52.42366065 73.25389697]
  [48.84498667 49.28798941 36.71220461]
  [46.32095657 47.7328842 18.58971283]
  [50.56297354 50.34647533 49.04732009]]

  [[58.33638332 51.93717508 74.24251925]
  [49.63206424 49.20100603 36.88896979]
  [52.36319954 47.06512954 19.94670506]
  [55.40388662 49.81148489 50.13451262]]]
```

1, 2, 3, 4 番目の合成変数を使った場合の予測値を求める。これは ncomp=4 を指定したときと同じ結果になる。

```
a = pls_predict(obj, newX=newX, comps=[1,2,3,4])
```

```
***** Predicted values
[[[60.16128321 53.12409669 78.652381 ]
  [46.00883968 46.51212623 34.48887842]
  [55.84530954 49.62322463 22.75170444]
  [59.0248579 52.54144757 51.71544973]]]
```

1, 4 番目の合成変数を使った場合の予測値を求める。

```
a = pls_predict(obj, newX=newX, comps=[1,4])
```

```
***** Predicted values
[[[56.67068701 54.43464486 75.44322706]
  [45.37589856 46.7384731 33.93658225]
  [49.80892427 50.29627556 21.38044077]
  [53.77726314 52.70873408 51.6190778 ]]
```

これ以降はスコア (合成変数の値) を求める (type="score")。

5 個の合成変数の値を求める。

```
a = pls_predict(obj, newX=newX, type="score")
```

```
***** Predicted scores
[[-25.6925055 11.40795704 -12.91503197 8.74765423 -1.47354701]
 [ 7.53410319 2.03973708 -2.18415926 -11.09023478 6.6877074 ]]
```

```
[ 24.01251194 15.65866891 -0.08300537 11.15059042 -11.80193958]
[-3.6819634 12.54538347 5.76279995 10.27567626 24.70246647]]
```

4 番目の合成変数の値を求める。

```
a = pls_predict(obj, newX=newX, ncomp=4, type="score")
```

```
***** Predicted scores
[[ 8.74765423]
 [-11.09023478]
 [ 11.15059042]
 [ 10.27567626]]
```

1 番目, 2 番目の合成変数の値を求める。

```
a = pls_predict(obj, newX=newX, ncomp=[1,2], type="score")
```

```
***** Predicted scores
[[-25.6925055 11.40795704]
 [ 7.53410319 2.03973708]
 [ 24.01251194 15.65866891]
 [ -3.6819634 12.54538347]]
```

1 番目, 2 番目, 3 番目, 4 番目の合成変数の値を求める。

```
a = pls_predict(obj, newX=newX, comps=[1,2,3,4], type="score")
```

```
***** Predicted scores
[[-25.6925055 11.40795704 -12.91503197 8.74765423]
 [ 7.53410319 2.03973708 -2.18415926 -11.09023478]
 [ 24.01251194 15.65866891 -0.08300537 11.15059042]
 [ -3.6819634 12.54538347 5.76279995 10.27567626]]
```

1 番目, 4 番目の合成変数の値を求める。

```
a = pls_predict(obj, newX=newX, comps=[1,4], type="score")
```

```
***** Predicted scores
[[-25.6925055 8.74765423]
 [ 7.53410319 -11.09023478]
 [ 24.01251194 11.15059042]
 [ -3.6819634 10.27567626]]
```