PCR, CCA, and other pattern-based regression techniques

Michael K. Tippett

International Research Institute for Climate and Society
The Earth Institute, Columbia University

Statistical Methods in Seasonal Prediction, ICTP
Aug 2-13, 2010

Principal component analysis and regression

Key idea of PCA: *data compression*

- Many (colinear) variables are replaced by a few new variables.
- The new variables are optimally chosen to approximate the original variables.

[Assumption: "important" components have large variance]

How is PCA useful in regression problems?

Pop quiz

- What is the variance of a PC (time-series)?
- What is the correlation between PCs?

Fact about regression

- (Easy) If $y = ax$ is regression between x and y, and x and y have unit variance, what does a measure?
- (Hard) Linear (invertible) transformations of the data transform the regression coefficients the same way.

$$y = Ax$$

$$y' = Ly \; , \qquad x' = Mx$$

$$y = Ax \rightarrow y' = Ly = LAx = LAM^{-1}Mx = (LAM^{-1})x'$$

(LAM^{-1}) = regression coefficient matrix between x' and y'

Pop quiz

- What is the variance of a PC (time-series)?
- What is the correlation between PCs?

Fact about regression

- (Easy) If $y = ax$ is regression between x and y, and x and y have unit variance, what does a measure?
- (Hard) Linear (invertible) transformations of the data transform the regression coefficients the same way.

$$y = Ax$$

$$y' = Ly \; , \qquad x' = Mx$$

$$y = Ax \rightarrow y' = Ly = LAx = LAM^{-1}Mx = (LAM^{-1})x'$$

(LAM^{-1}) = regression coefficient matrix between x' and y'

Pop quiz

- What is the variance of a PC (time-series)?
- What is the correlation between PCs?

Fact about regression

- (Easy) If $y = ax$ is regression between x and y, and x and y have unit variance, what does a measure?
- (Hard) Linear (invertible) transformations of the data transform the regression coefficients the same way.

$$y = Ax$$

$$y' = Ly\,, \qquad x' = Mx$$

$$y = Ax \rightarrow y' = Ly = LAx = LAM^{-1}Mx = (LAM^{-1})x'$$

(LAM^{-1}) = regression coefficient matrix between x' and y'

Pop quiz

- What is the variance of a PC (time-series)?
- What is the correlation between PCs?

Fact about regression

- (Easy) If $y = ax$ is regression between x and y, and x and y have unit variance, what does a measure?
- (Hard) Linear (invertible) transformations of the data transform the regression coefficients the same way.

$$y = Ax$$

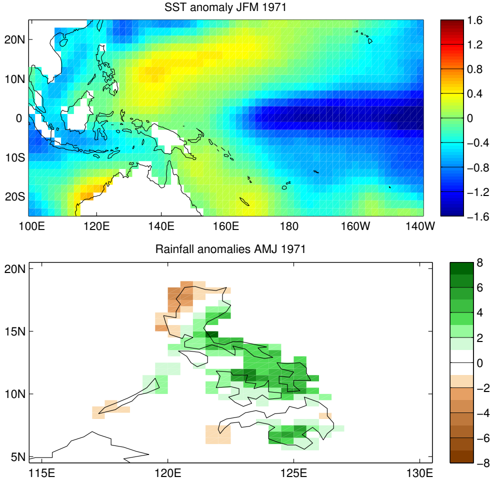
$$y' = Ly\ , \qquad x' = Mx$$

$$y = Ax \rightarrow y' = Ly = LAx = LAM^{-1}Mx = (LAM^{-1})x'$$

(LAM^{-1}) = regression coefficient matrix between x' and y'

Example: Philippines

Problem: predict gridded April-June precipitation over the Philippines from proceeding (January-March) SST.



Example: Philippines

Problem: predict gridded April-June precipitation over the Philippines from proceeding (January-March) sea surface temperature.

Details:

- Data from 1971-2007 (37 years).
- 194 precipitation gridpoints.
- 1378 SST gridpoints.

What is the problem?

PCA and regression

For climate forecasts, the length of the historical record severely limits the number of predictors

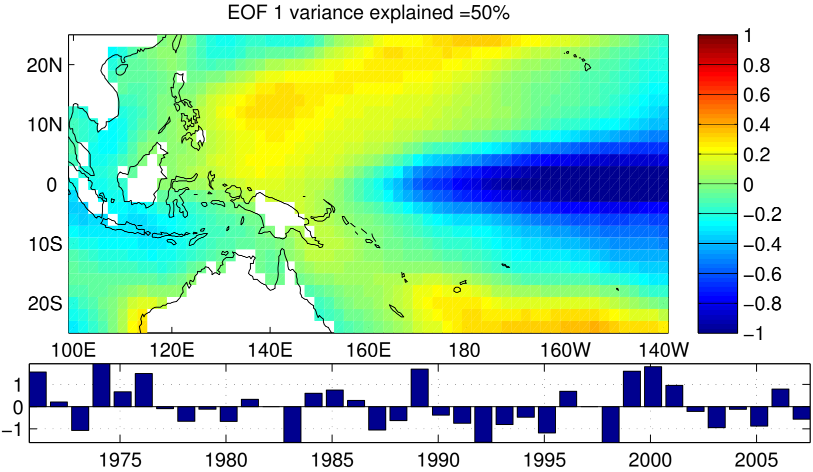
If the predictors are spatial fields such as SST or the output of a GCM, the number of grid point values (100's, 1000's) is large compared to the number time samples (10's for climate)

Need to represent the information in the predictor spatial field using fewer numbers.

- Spatial averages e.g., NINO 3.4.
- Principal component analysis (PCA).
  - Weighted spatial average.
  - Weights are chosen in an optimal manner to maximize explained variance.

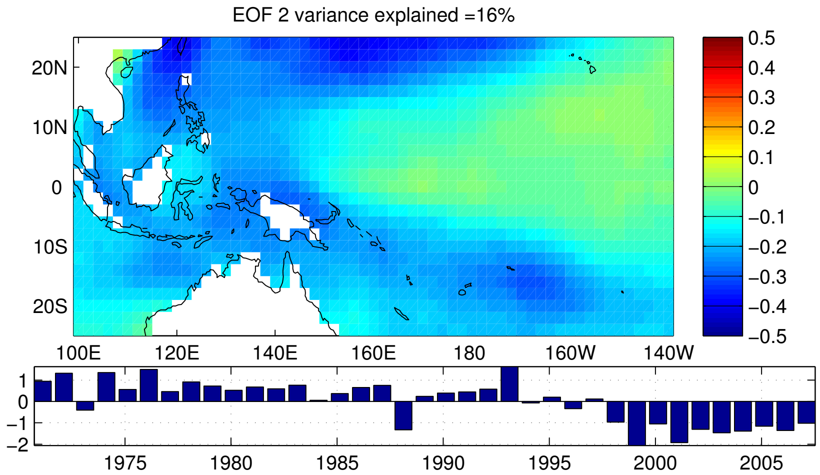
Example: PCA of SST

EOF 1 – Correlation with NINO 3.4 = -0.96



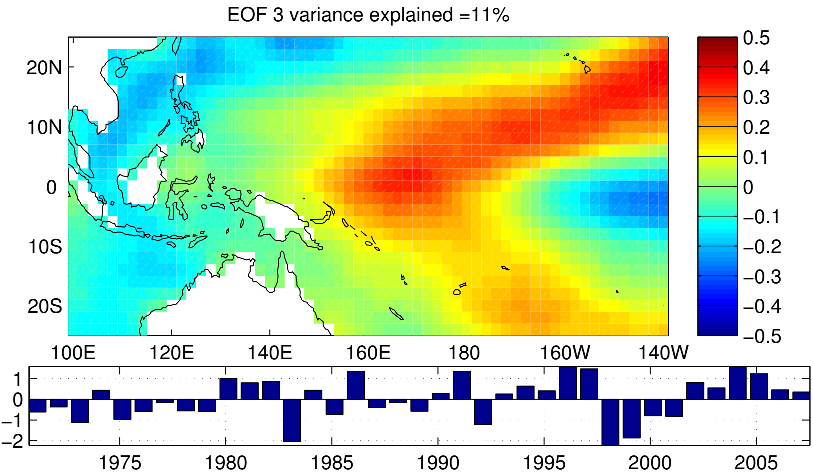
Example: PCA of SST

EOF 2



Example: PCA of SST

EOF 3



Principal component regression

PCR

- $\hat{y} = a_1 x_1 + a_2 x_2 + \ldots a_m x_m + b$
- Predictors x_i are PCs.

In this example:

- y = observed precipitation at a gridpoint.
- PCs of SST anomalies.

How many PCs to use?

Predictor selection problem.

Principal component regression

PCR

- $\hat{y} = a_1 x_1 + a_2 x_2 + \ldots a_m x_m + b$
- Predictors x_i are PCs.

In this example:

- y = observed precipitation at a gridpoint.
- PCs of SST anomalies.

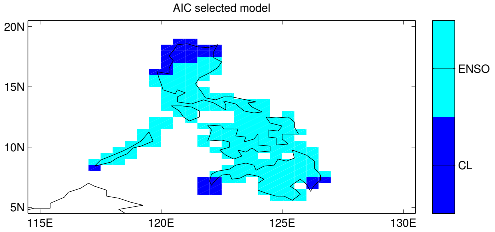
How many PCs to use?

Predictor selection problem.

Example: Philippines

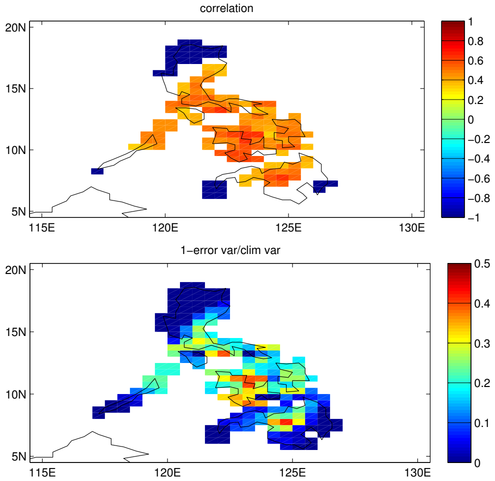
Two models: climatology or ENSO PC as predictor.

Use AIC to select model.



Example: Philippines

- This model seems to have some skill (cross-validated)
- Why the negative correlation? [Later]
- How are the two skill measures related in-sample?



PCA and regression

Is there any benefit to using PCA on the predictand as well as the predictors?

Is there any benefit to predicting the PCs of y rather than y?

Perhaps. One could imagine a spatial average (like a PC) being more predictable than a value at a gridpoint.

PCA and regression

Is there any benefit to using PCA on the predictand as well as the predictors?

Is there any benefit to predicting the PCs of y rather than y?

Perhaps. One could imagine a spatial average (like a PC) being more predictable than a value at a gridpoint.

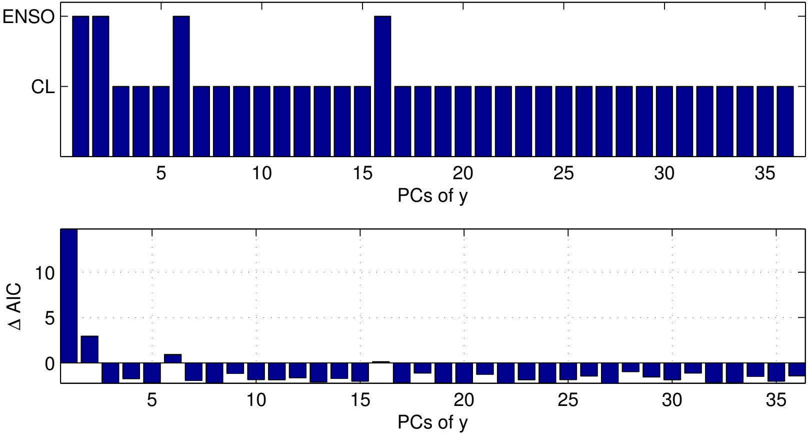
PCA and regression

- Predicting the PCs of y leads to a different predictor selection problem
- Before: select a model for each gridpoint?
- Now: select a model for each PC?

Example: Philippines

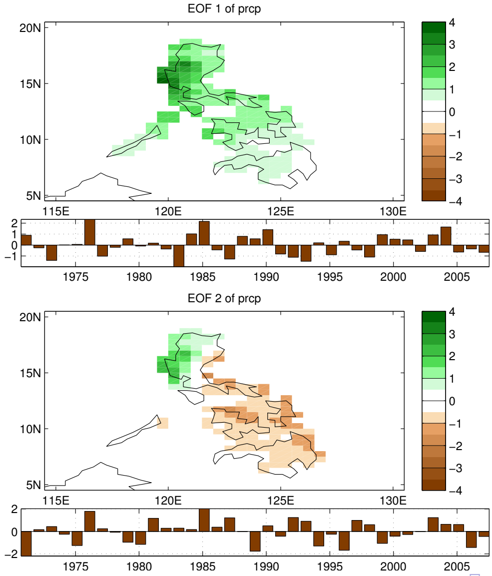
36 PCs of y. (Why?)

Use AIC to select model. ENSO or climatology.



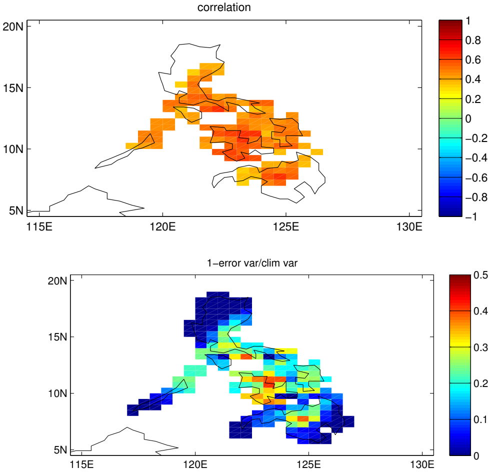
Example: Philippines

First 2 EOFs of AMJ precipitation:



Example: Philippines

- Correlations of gridpoint and pattern regressions are similar.
- Normalized error of gridpoint and pattern regressions are similar.

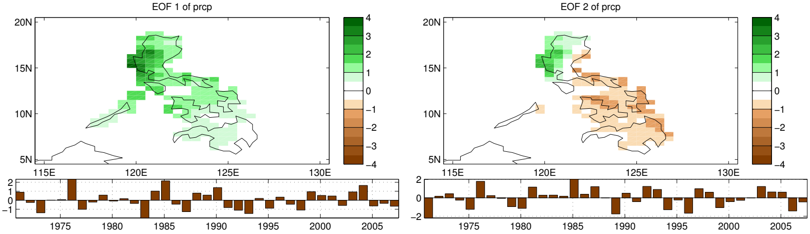


Example: Philippines

Pattern regression is:

$$\hat{PC}_{y1} = 0.61\ PC_{x1}$$
$$\hat{PC}_{y2} = -0.36\ PC_{x1}$$

What do these numbers mean? (Hint: PCs have unit variance.)



Example: Philippines

Reconstructing the spatial field:

$$\text{Predicted rainfall} = \text{Climatology} + \hat{PC}_{y1}\text{EOF}_{y1} + \hat{PC}_{y2}\text{EOF}_{y2}$$

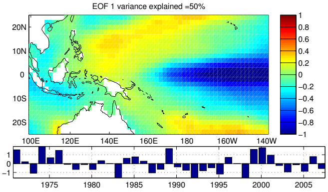
Difference between prediction and climatology (anomaly) is:

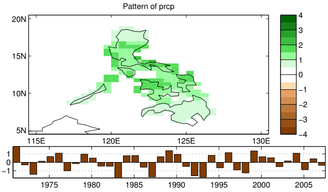
$$\begin{aligned} &\text{Predicted anomaly} = \\ \hat{PC}_{y1}\text{EOF}_{y1} + \hat{PC}_{y2}\text{EOF}_{y2} &= 0.61\ \text{PC}_{x1}\text{EOF}_{y1} - 0.36\ \text{PC}_{x1}\text{EOF}_{y1} \\ &= \text{PC}_{x1}(0.61\ \text{EOF}_{y1} - 0.36\ \text{EOF}_{y1}) \end{aligned}$$

Is this simpler? Why?

Example: Philippines

One pattern of rainfall goes with one pattern of SST.





What is the time series of the pattern?

Example: Philippines

Define the pattern to be

$$P \equiv \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1})$$

(Why this scaling of P?) The time series of the pattern P is:

$$\text{TS} = \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ PC}_{y1} - 0.36 \text{ PC}_{y1})$$

What is the variance of TS? (Hint: PCs are independent.) Key

$$\begin{aligned} \text{Predicted anomaly} &= \text{PC}_{x1}(0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1}) \\ &= 0.71 \text{ PC}_{x1} P \end{aligned}$$

$$\hat{\text{TS}} = 0.71 \text{ PC}_{x1}$$

What is 0.71? Hint: TS and PC_{x1} have unit variance.

Example: Philippines

Define the pattern to be

$$P \equiv \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1})$$

(Why this scaling of P?) The time series of the pattern P is:

$$\text{TS} = \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ PC}_{y1} - 0.36 \text{ PC}_{y1})$$

What is the variance of TS? (Hint: PCs are independent.) Key

$$\begin{aligned} \text{Predicted anomaly} &= \text{PC}_{x1}(0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1}) \\ &= 0.71 \text{ PC}_{x1} P \end{aligned}$$

$$\hat{\text{TS}} = 0.71 \text{ PC}_{x1}$$

What is 0.71? Hint: TS and PC_{x1} have unit variance.

Example: Philippines

Define the pattern to be

$$P \equiv \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1})$$

(Why this scaling of P?) The time series of the pattern P is:

$$\text{TS} = \frac{1}{\sqrt{0.61^2 + 0.36^2}} (0.61 \text{ PC}_{y1} - 0.36 \text{ PC}_{y1})$$

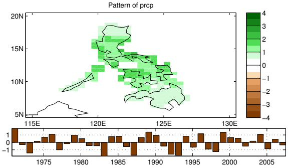
What is the variance of TS? (Hint: PCs are independent.) Key

$$\begin{aligned} \text{Predicted anomaly} &= \text{PC}_{x1}(0.61 \text{ EOF}_{y1} - 0.36 \text{ EOF}_{y1}) \\ &= 0.71 \text{ PC}_{x1} P \end{aligned}$$

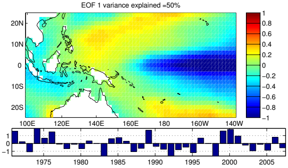
$$\hat{\text{TS}} = 0.71 \text{ PC}_{x1}$$

What is 0.71? Hint: TS and PC_{x1} have unit variance.

In summary,



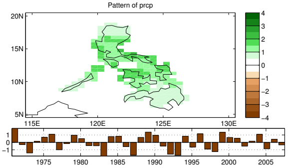
$= 0.71 \times$



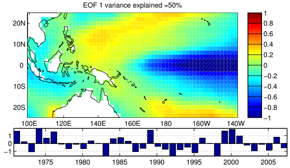
In general, any pattern regression can be decomposed into pairs of patterns related by the correlation of their time series.

Canonical correlation analysis (CCA) is an example of such a decomposition.

In summary,



$= 0.71\times$



In general, any pattern regression can be decomposed into pairs of patterns related by the correlation of their time series.

Canonical correlation analysis (CCA) is an example of such a decomposition.

Pattern regression

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_l \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1m} \\ a_{21} & a_{22} & . & . & . & a_{2m} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{l1} & a_{l2} & . & . & . & a_{lm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{bmatrix}$$

- l predictand PCs
- m predictor PCs
- $l \times m$ regression coefficients.

$$A = \mathrm{Cov}\left(\mathrm{PC}_y, \mathrm{PC}_x\right) \left[\mathrm{Cov}\left(\mathrm{PC}_x, \mathrm{PC}_x^T\right)\right]^{-1}$$

Pattern regression

$$y = Ax$$

$$A = \mathrm{Cov}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)\left[\mathrm{Cov}\left(\mathrm{PC}_x, \mathrm{PC}_x^T\right)\right]^{-1}$$

What is $\mathrm{Cov}\left(\mathrm{PC}_x, \mathrm{PC}_x^T\right)$? Why?
Hint: PCs are . . . .

$$A = \mathrm{Cov}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)$$

What do the elements of A measure?

$$A = \mathrm{Corr}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)$$

Pattern regression

$$y = Ax$$

$$A = \mathrm{Cov}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)\left[\mathrm{Cov}\left(\mathrm{PC}_x, \mathrm{PC}_x^T\right)\right]^{-1}$$

What is $\mathrm{Cov}\left(\mathrm{PC}_x, \mathrm{PC}_x^T\right)$? Why?
Hint: PCs are ....

$$A = \mathrm{Cov}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)$$

What do the elements of A measure?

$$A = \mathrm{Corr}\left(\mathrm{PC}_y, \mathrm{PC}_x\right)$$

Pattern regression

$$y = Ax$$

$$A = \begin{bmatrix} \mathrm{Corr}(PC_{y1}, PC_{x1}) & \mathrm{Corr}(PC_{y1}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{y1}, PC_{xm}) \\ \mathrm{Corr}(PC_{y2}, PC_{x1}) & \mathrm{Corr}(PC_{y2}, PC_{y2}) & \cdots & \mathrm{Corr}(PC_{y2}, PC_{xm}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Corr}(PC_{yl}, PC_{x1}) & \mathrm{Corr}(PC_{yl}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{yl}, PC_{xm}) \end{bmatrix}$$

In general, each predicted PC of y depends on *all* the PCs of x.

What if A were diagonal? Is it likely that A is diagonal?

Angle.

Pattern regression

$$y = Ax$$

$$A = \begin{bmatrix} \mathrm{Corr}(PC_{y1}, PC_{x1}) & \mathrm{Corr}(PC_{y1}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{y1}, PC_{xm}) \\ \mathrm{Corr}(PC_{y2}, PC_{x1}) & \mathrm{Corr}(PC_{y2}, PC_{y2}) & \cdots & \mathrm{Corr}(PC_{y2}, PC_{xm}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Corr}(PC_{yl}, PC_{x1}) & \mathrm{Corr}(PC_{yl}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{yl}, PC_{xm}) \end{bmatrix}$$

In general, each predicted PC of y depends on *all* the PCs of x.

What if A were diagonal? Is it likely that A is diagonal?

Angle.

Pattern regression

$$y = Ax$$

$$A = \begin{bmatrix} \mathrm{Corr}(PC_{y1}, PC_{x1}) & \mathrm{Corr}(PC_{y1}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{y1}, PC_{xm}) \\ \mathrm{Corr}(PC_{y2}, PC_{x1}) & \mathrm{Corr}(PC_{y2}, PC_{y2}) & \cdots & \mathrm{Corr}(PC_{y2}, PC_{xm}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Corr}(PC_{yl}, PC_{x1}) & \mathrm{Corr}(PC_{yl}, PC_{x2}) & \cdots & \mathrm{Corr}(PC_{yl}, PC_{xm}) \end{bmatrix}$$

In general, each predicted PC of y depends on *all* the PCs of x.

What if A were diagonal? Is it likely that A is diagonal?

Angle.

Pattern regression

$$y = Ax$$

$$A = \begin{bmatrix} \mathrm{Corr}(\mathrm{PC}_{y1}, \mathrm{PC}_{x1}) & \mathrm{Corr}(\mathrm{PC}_{y1}, \mathrm{PC}_{x2}) & . & . & . & \mathrm{Corr}(\mathrm{PC}_{y1}, \mathrm{PC}_{xm}) \\ \mathrm{Corr}(\mathrm{PC}_{y2}, \mathrm{PC}_{x1}) & \mathrm{Corr}(\mathrm{PC}_{y2}, \mathrm{PC}_{y2}) & . & . & . & \mathrm{Corr}(\mathrm{PC}_{y2}, \mathrm{PC}_{xm}) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathrm{Corr}(\mathrm{PC}_{yl}, \mathrm{PC}_{x1}) & \mathrm{Corr}(\mathrm{PC}_{yl}, \mathrm{PC}_{x2}) & . & . & . & \mathrm{Corr}(\mathrm{PC}_{yl}, \mathrm{PC}_{xm}) \end{bmatrix}$$

In general, each predicted PC of y depends on *all* the PCs of x.

What if A were diagonal? Is it likely that A is diagonal?

Angle.

Pattern regression

- To decompose the regression into pairs of patterns, diagonalize A.
- Many ways to diagonalize A. The singular value decomposition (SVD) is:

$$A = USV^T$$

where U and S are orthogonal and S is diagonal.

(orthogonal matrix = columns are unit vectors = preserves angles and magnitudes)

Substituting

$$y = Ax = USV^T x$$

or

$$y' = Sx'$$

where $y' = U^T y$ and $x' = V^T x$

Pattern regression

- To decompose the regression into pairs of patterns, diagonalize A.
- Many ways to diagonalize A. The singular value decomposition (SVD) is:

  $$A = USV^T$$

  where U and S are orthogonal and S is diagonal.

  (orthogonal matrix = columns are unit vectors = preserves angles and magnitudes)

Substituting

$$y = Ax = USV^T x$$

or

$$y' = Sx'$$

where $$y' = U^T y \text{ and } x' = V^T x$$

Diagonalized pattern regression

What can we say about the new variables

$$y' = U^T y \text{ and } x' = V^T x?$$

- y' and x' have unit variance and are uncorrelated (like PCs).
  - PCs have unit variance and are uncorrelated.
  - Orthogonal transformation of PCs gives new uncorrelated (angle) variables with unit variance (magnitude).
- Each new predictand related to just one new predictor.
  - $y' = Sx'$ S is diagonal,
- What do the values of S measure?

  (Hint: what is the regression coefficient for two variables with unit variance?)

Diagonalized regression and CCA

This procedure is the same as *canonical correlation analysis*.

- Regress PCs (uncorrelated unit variance) of y and y.

$$y = Ax$$

- Use SVD of A to get diagonal relation: $y' = Sx'$.
  - New variables (canonical variates) are linear (orthogonal) combinations of the PCs.
  - New variables have unit variance and are uncorrelated.
  - Associated patterns are linear combinations of EOFs. (Generally not orthogonal).
  - Elements of S are correlations (canonical correlations).

More CCA

CCA is usually described as finding linear combinations of the x's and the y's which have maximum correlation.

Did we do that???

Finding maximum correlation between linear combinations of x and y is the same as finding maximum correlation between linear combinations of x' and y'. Why?

This means we can look at the correlation between linear combinations of x' and y'.

More CCA

CCA is usually described as finding linear combinations of the x's and the y's which have maximum correlation.

Did we do that???

Finding maximum correlation between linear combinations of x and y is the same as finding maximum correlation between linear combinations of x' and y'. Why?

This means we can look at the correlation between linear combinations of x' and y'.

A calculation

$$\text{Corr}\left(\sum a_i x_i', \sum b_j y_j'\right) = \frac{\text{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right)}{\sqrt{\text{Var}\left(\sum a_i x_i'\right)\text{Var}\left(\sum b_j y_j'\right)}}$$

$$\text{Var}\left(\sum a_i x_i'\right) = \sum a_i^2 \,\text{Var}\left(x_i'\right) = \sum a_i^2 = \|a\|^2$$

$$\text{Var}\left(\sum b_j y_j'\right) = \sum b_j^2 \,\text{Var}\left(y_j'\right) = \sum b_j^2 = \|b\|^2$$

$$\begin{aligned}\text{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right) &= \sum a_i b_i \text{Cov}\left(x_i', y_i'\right) \\ &= \sum a_i b_i S_i \le S_1 \sum a_i b_i \le \|a\|\,\|b\|\end{aligned}$$

$$\text{Corr}\left(\sum a_i x_i', \sum b_i y_i'\right) \le S_1$$

A calculation

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_j y_j'\right) = \frac{\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right)}{\sqrt{\mathrm{Var}\left(\sum a_i x_i'\right)\mathrm{Var}\left(\sum b_j y_j'\right)}}$$

$$\mathrm{Var}\left(\sum a_i x_i'\right) = \sum a_i^2 \,\mathrm{Var}\left(x_i'\right) = \sum a_i^2 = \|a\|^2$$

$$\mathrm{Var}\left(\sum b_j y_j'\right) = \sum b_j^2 \,\mathrm{Var}\left(y_j'\right) = \sum b_j^2 = \|b\|^2$$

$$\begin{aligned}\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right) &= \sum a_i b_i \mathrm{Cov}\left(x_i', y_i'\right)\\ &= \sum a_i b_i S_i \le S_1 \sum a_i b_i \le \|a\|\,\|b\|\end{aligned}$$

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_i y_i'\right) \le S_1$$

A calculation

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_j y_j'\right) = \frac{\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right)}{\sqrt{\mathrm{Var}\left(\sum a_i x_i'\right)\,\mathrm{Var}\left(\sum b_j y_j'\right)}}$$

$$\mathrm{Var}\left(\sum a_i x_i'\right) = \sum a_i^2\,\mathrm{Var}\left(x_i'\right) = \sum a_i^2 = \|a\|^2$$

$$\mathrm{Var}\left(\sum b_j y_j'\right) = \sum b_j^2\,\mathrm{Var}\left(y_j'\right) = \sum b_j^2 = \|b\|^2$$

$$\begin{aligned}\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right) &= \sum a_i b_i \mathrm{Cov}\left(x_i', y_i'\right)\\ &= \sum a_i b_i S_i \leq S_1 \sum a_i b_i \leq \|a\|\,\|b\|\end{aligned}$$

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_i y_i'\right) \leq S_1$$

A calculation

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_j y_j'\right) = \frac{\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right)}{\sqrt{\mathrm{Var}\left(\sum a_i x_i'\right)\mathrm{Var}\left(\sum b_j y_j'\right)}}$$

$$\mathrm{Var}\left(\sum a_i x_i'\right) = \sum a_i^2\,\mathrm{Var}\left(x_i'\right) = \sum a_i^2 = \|a\|^2$$

$$\mathrm{Var}\left(\sum b_j y_j'\right) = \sum b_j^2\,\mathrm{Var}\left(y_j'\right) = \sum b_j^2 = \|b\|^2$$

$$\begin{aligned}\mathrm{Cov}\left(\sum a_i x_i', \sum b_j y_j'\right) &= \sum a_i b_i \mathrm{Cov}\left(x_i', y_i'\right)\\ &= \sum a_i b_i S_i \le S_1 \sum a_i b_i \le \|a\|\,\|b\|\end{aligned}$$

$$\mathrm{Corr}\left(\sum a_i x_i', \sum b_i y_i'\right) \le S_1$$

More CCA components

Look for the linear combination of x and y that maximizes correlation *but* is uncorrelated with the first component means

- looking for the linear combination of x_i' and y_i' $i = 2, 3, \ldots$ that maximizes correlation. Why?
- Previous argument give that it is S_2.

CCA and regression

Knowing CCA is regression is useful ...

- What happens if many (compared to sample size) PCs are included in a CCA calculation? What happens to the canonical correlations?
- How can the number of PCs included in CCA be decided?

CCA and regression

Knowing CCA is regression is useful ...

- What happens if many (compared to sample size) PCs are included in a CCA calculation? What happens to the canonical correlations?
- How can the number of PCs included in CCA be decided?

CCA and regression

Knowing CCA is regression is useful ...

- What happens if many (compared to sample size) PCs are included in a CCA calculation? What happens to the canonical correlations?
- How can the number of PCs included in CCA be decided?

Other pattern regression methods

Other diagonalizations of the regression coefficient matrix A (based on variants of the SVD) give other diagonalized pattern regressions with components that optimize other quantities.

E.g., Redundancy analysis give components that maximize explained variance.

Maximum covariance analysis (MCA) finds components with maximum covariance. *However*, the regression between these patterns is generally not diagonal–no simple relations between pairs of patterns.

Summary

- PCA compresses data and is useful in regressions. In PCR, PCs are the predictors.
- It can be useful to use PCs as predictands, too.
- Diagonalizing regressions between PCs decomposes the regression in pairs of patterns.
- CCA diagonalizes the regression and find the components with maximum correlation.