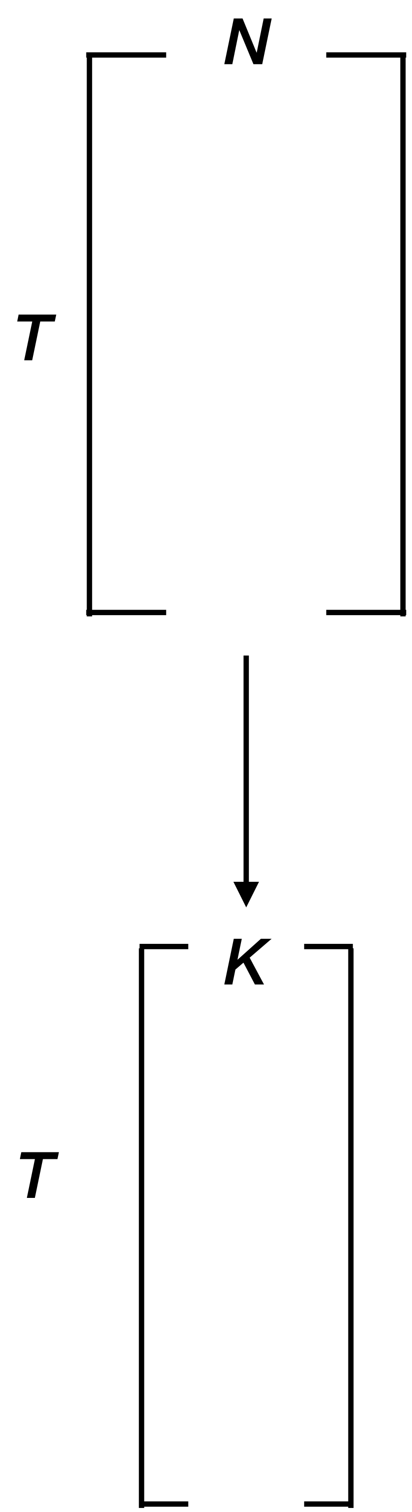
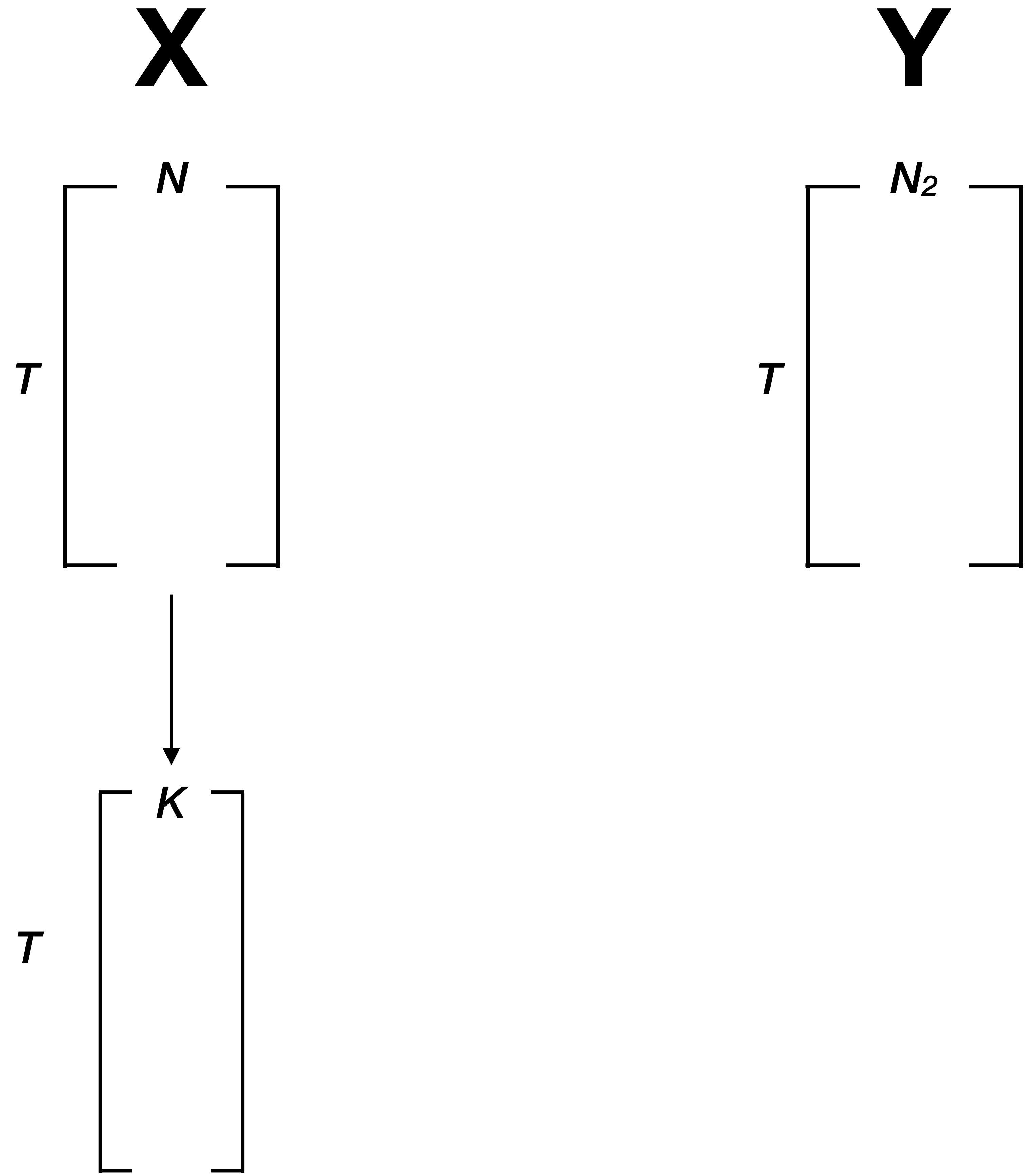
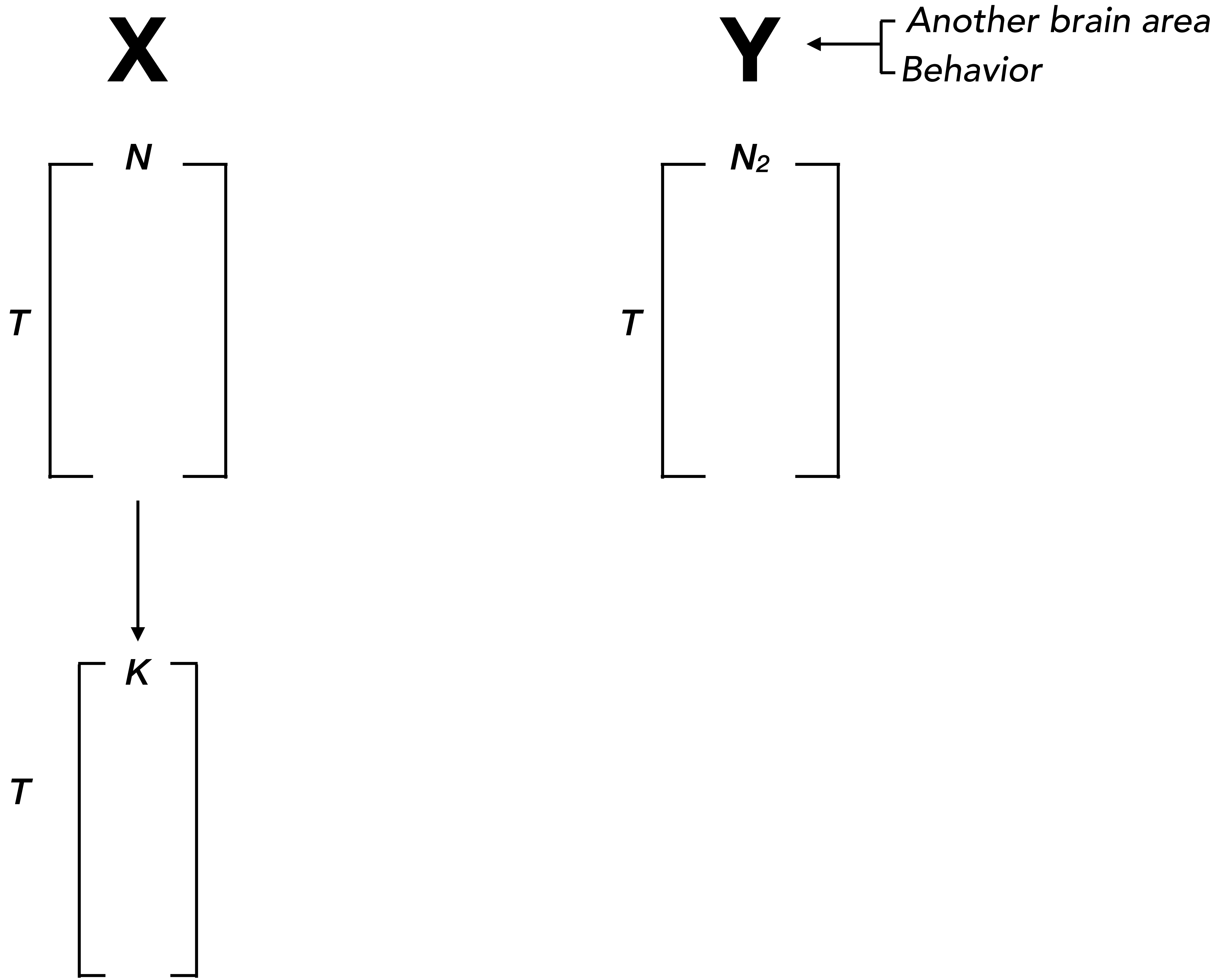
Linear dimensionality reduction across multiple datasets

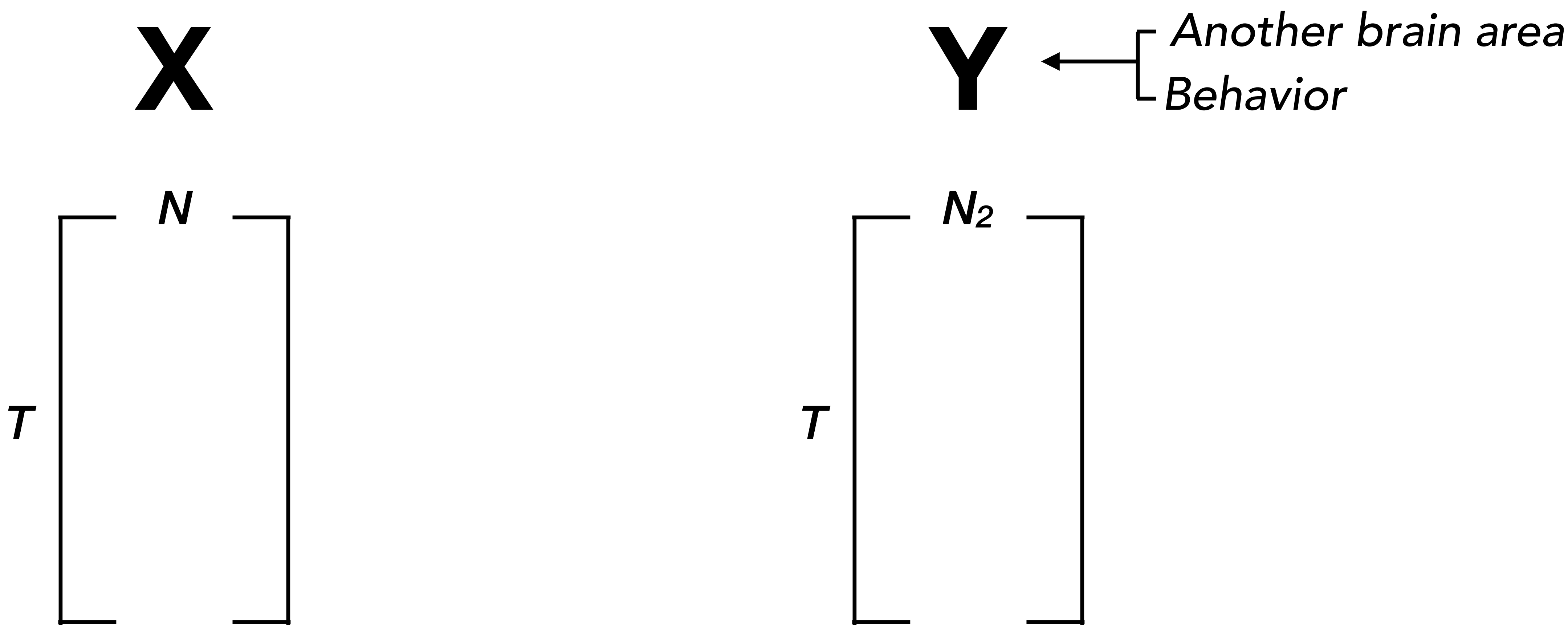
Josh Glaser
Advanced Theory Seminar

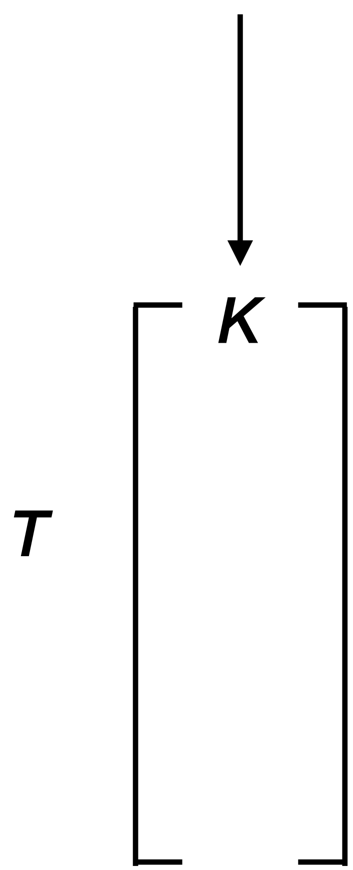
X




X
Y
N
N_2
T
T
K
T


X
Y
Another brain area
Behavior
N
T
N_2
T
K
T





- *Reduce noise*
- *Visualization*
- *Science!*

Overview

1. Reducing dimensionality in one space
   - PCA and regression refreshers
   - Principal components regression
   - Reduced rank regression
   - Demixed PCA
2. Reducing dimensionality in two spaces
   - Canonical correlation analysis
   - Partial least squares

Principal Components Analysis

Principal Components Analysis

- Find orthogonal linear low-D transformation that maximizes variance.

$$\underset{w \in \mathbb{R}^K}{\arg\max} \left(\| \mathbf{Xw} \|^2 \right) \textbf{ s.t. } \| \mathbf{w} \| = \mathbf{I}$$

Principal Components Analysis

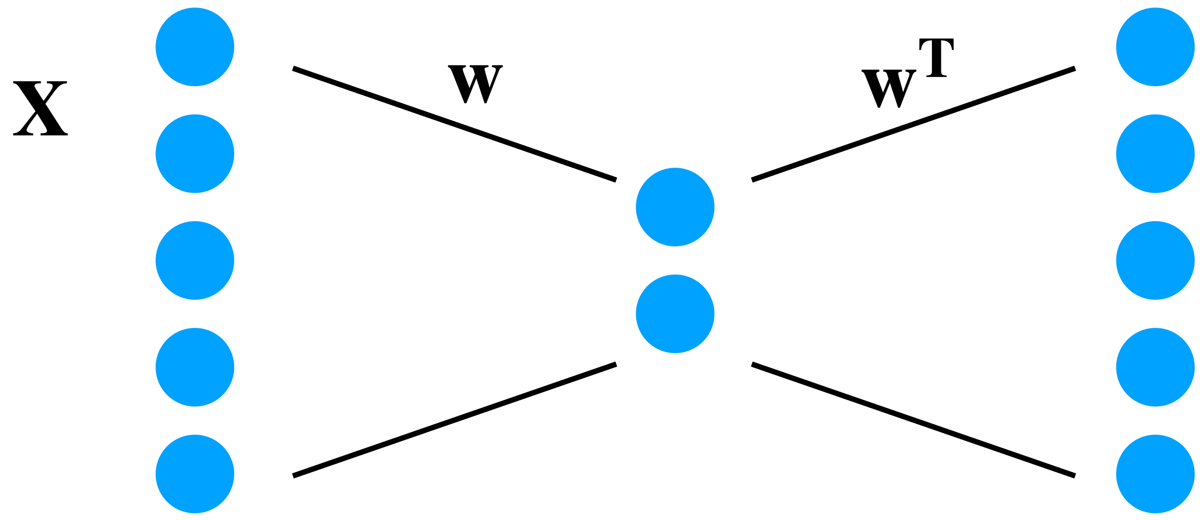
- Find orthogonal linear low-D transformation that maximizes variance.

$$\underset{w \in \mathbb{R}^K}{\arg\max} \left(\| \mathbf{Xw} \|^2 \right) \textbf{ s.t. } \| \mathbf{w} \| = \mathbf{I}$$

- Find transformation that minimizes reconstruction error.

$$\underset{w \in \mathbb{R}^K}{\arg\min} \left(\| \mathbf{X} - \mathbf{Xww^T} \|^2 \right) \textbf{ s.t. } \| \mathbf{w} \| = \mathbf{I}$$

Principal Components Analysis

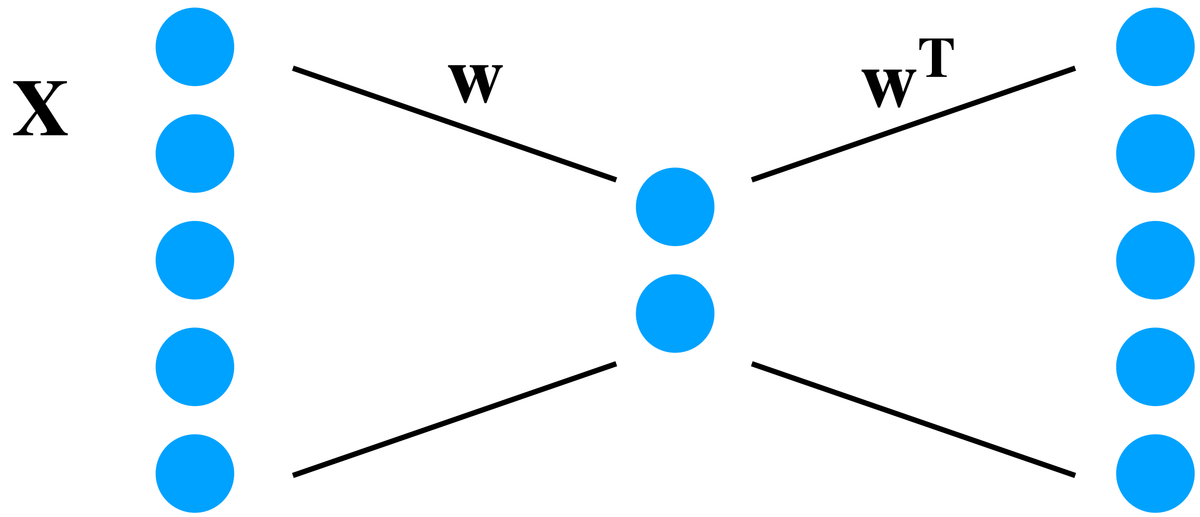


$$\mathbf{X_K} = \mathbf{Xww^T}$$

- Find transformation that minimizes reconstruction error.

$$\underset{w \in \mathbb{R}^K}{\arg\min} \left(\| \mathbf{X} - \mathbf{Xww^T} \|^2 \right) \textbf{ s.t. } \| \mathbf{w} \| = \mathbf{I}$$

Principal Components Analysis



$$\mathbf{X_K} = \mathbf{X}\mathbf{w}\mathbf{w}^{\mathbf{T}}$$

- Find transformation that minimizes reconstruction error.

$$\underset{w \in \mathbb{R}^K}{\arg\min} \left(\| \mathbf{X} - \mathbf{X}\mathbf{w}\mathbf{w}^{\mathbf{T}} \|^2 \right) \textbf{ s.t. } \| \mathbf{w} \| = \mathbf{I}$$

- $\mathbf{X_K}$ is the best rank K approximation (in terms of minimizing norm of reconstruction error) when solving with singular value decomposition (Eckart-Young theorem, 1936).

(OLS) Regression

$$\mathbf{X} \qquad \begin{array}{cc} & N \\ T & \left[\begin{array}{c} \quad \\ \quad \\ \quad \end{array}\right] \end{array} \qquad\qquad \mathbf{Y} \qquad \begin{array}{cc} & N_2 \\ T & \left[\begin{array}{c} \quad \\ \quad \\ \quad \end{array}\right] \end{array}$$

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|^{2} \right)$$

(OLS) Regression

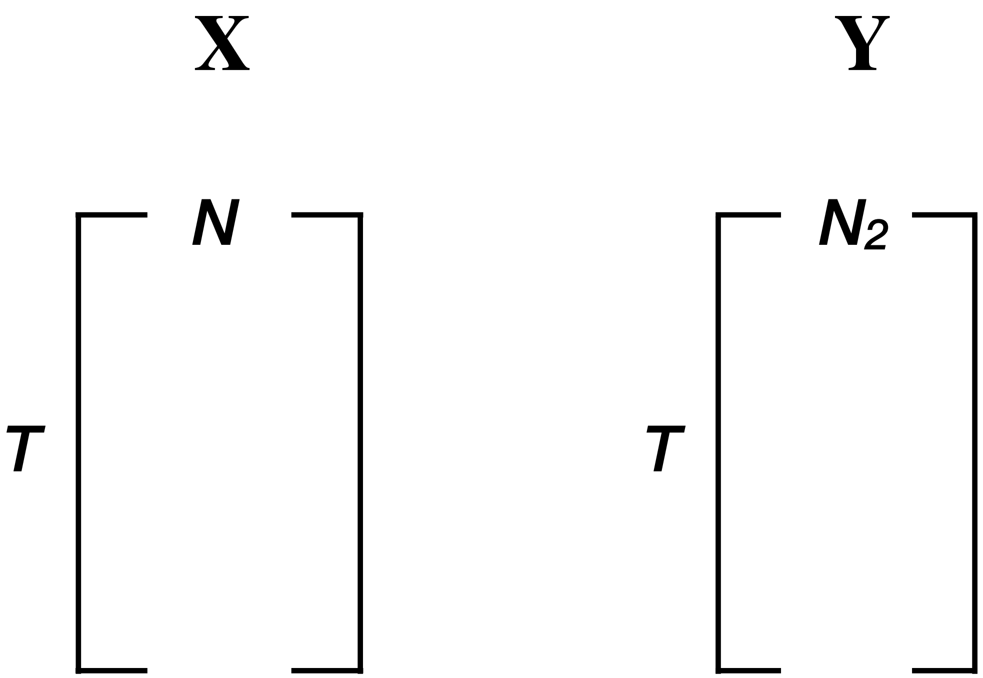
$$\mathbf{X} \qquad\qquad \mathbf{Y}$$

$$\begin{array}{cc} & N \\ T & \begin{bmatrix} \quad & \quad \\ & \\ & \\ & \end{bmatrix} \end{array} \qquad \begin{array}{cc} & N_2 \\ T & \begin{bmatrix} \quad & \quad \\ & \\ & \\ & \end{bmatrix} \end{array}$$

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|^{2} \right)$$

$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}} \mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

(OLS) Regression

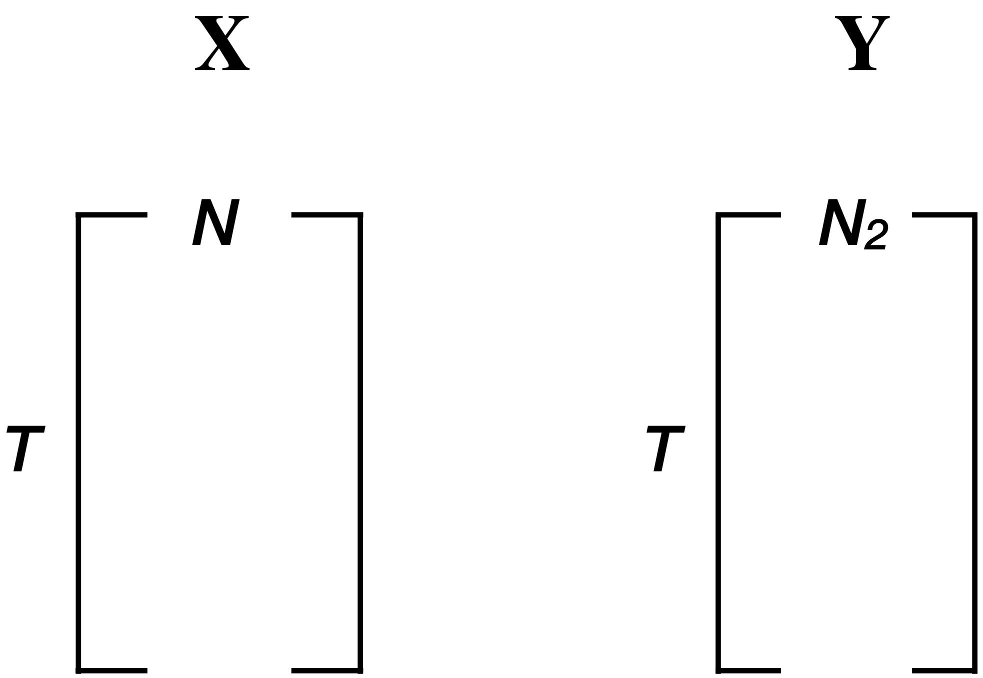


$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|^2 \right)$$

$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}} \mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

- This is equivalent to N_2 separate regressions (each for a separate column of **Y** to get a separate column of β).

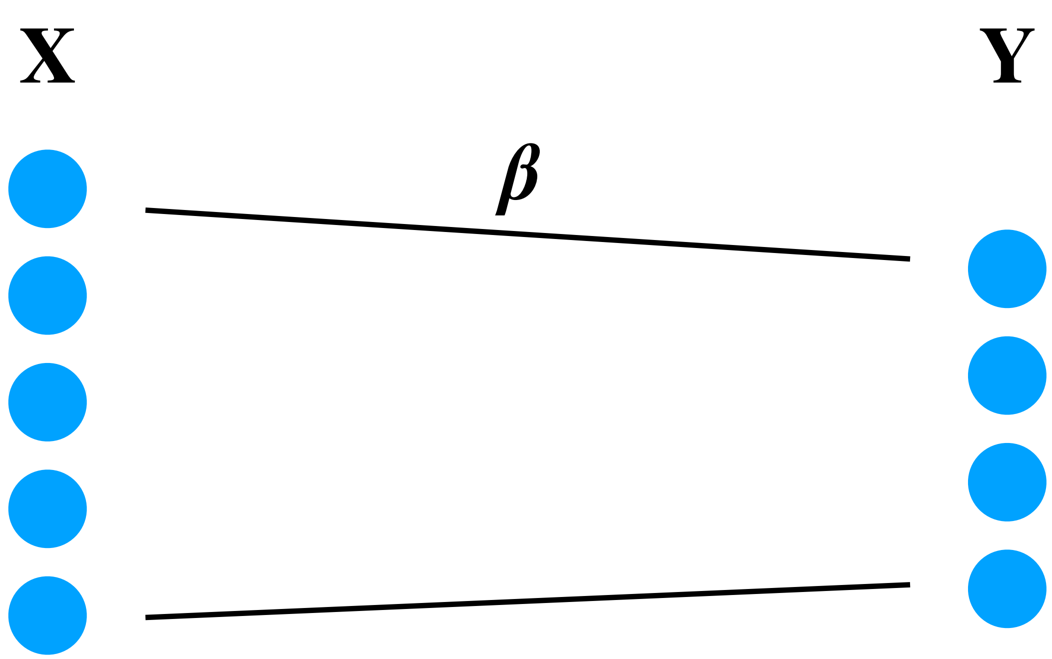
(OLS) Regression



$$\underset{\beta \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|^2 \right)$$

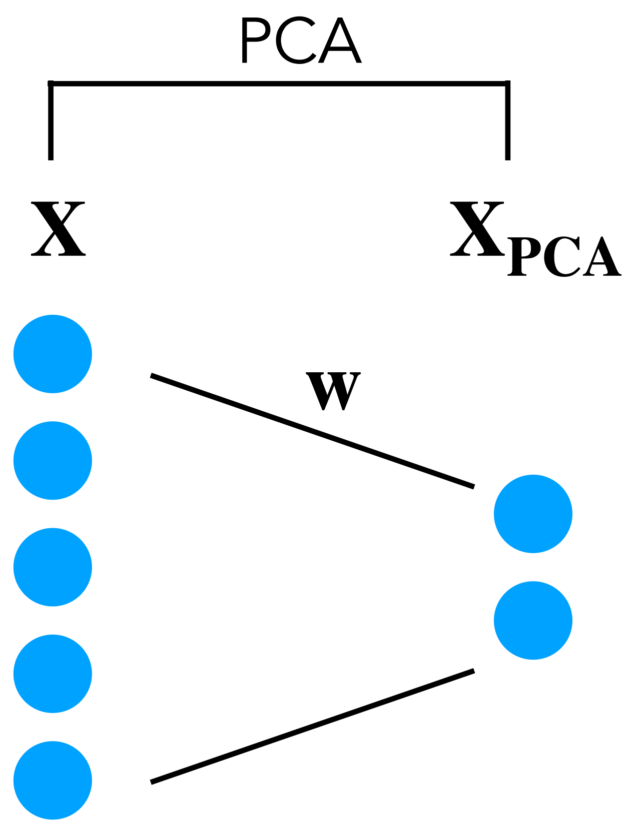
$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

- This is equivalent to N_2 separate regressions (each for a separate column of **Y** to get a separate column of β).



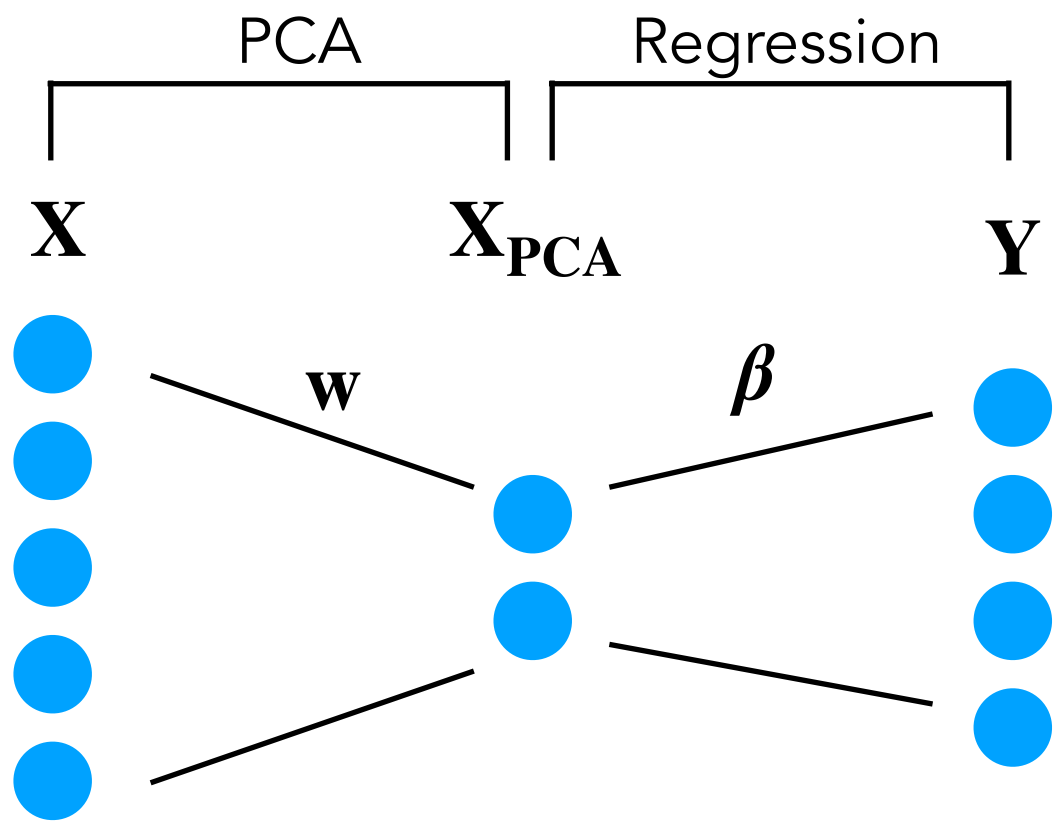
Principal Components Regression

- PCA + Regression



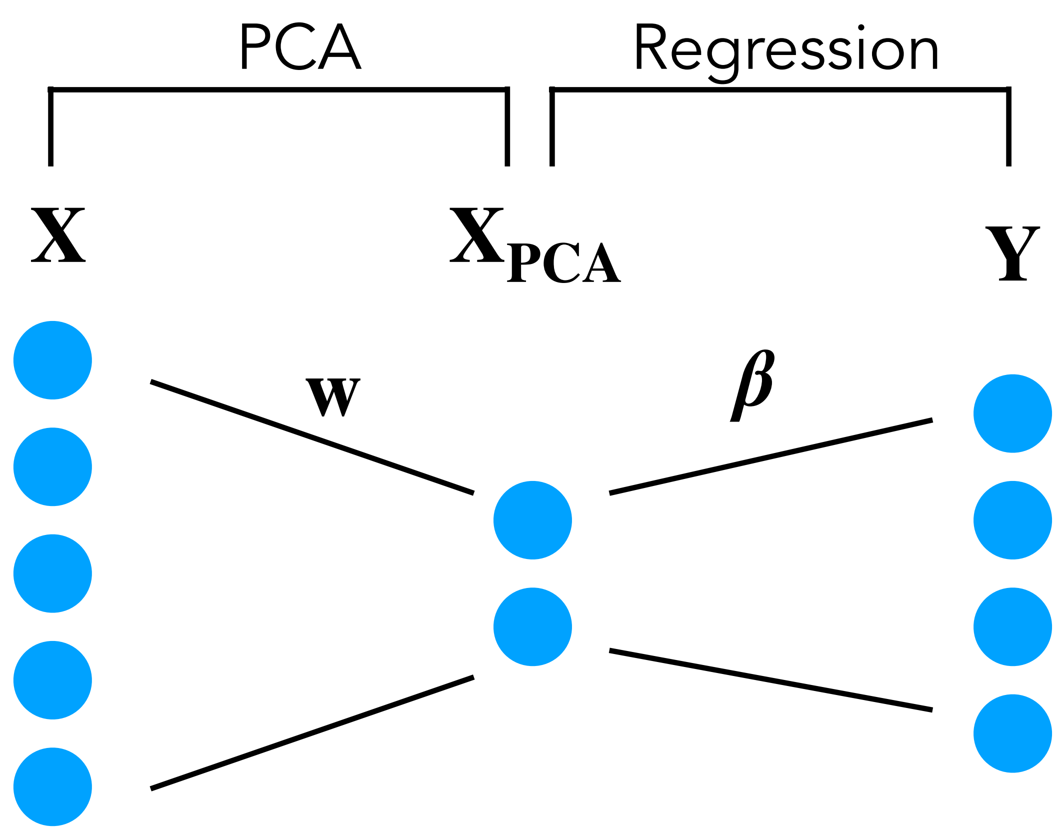
Principal Components Regression

- PCA + Regression



Principal Components Regression

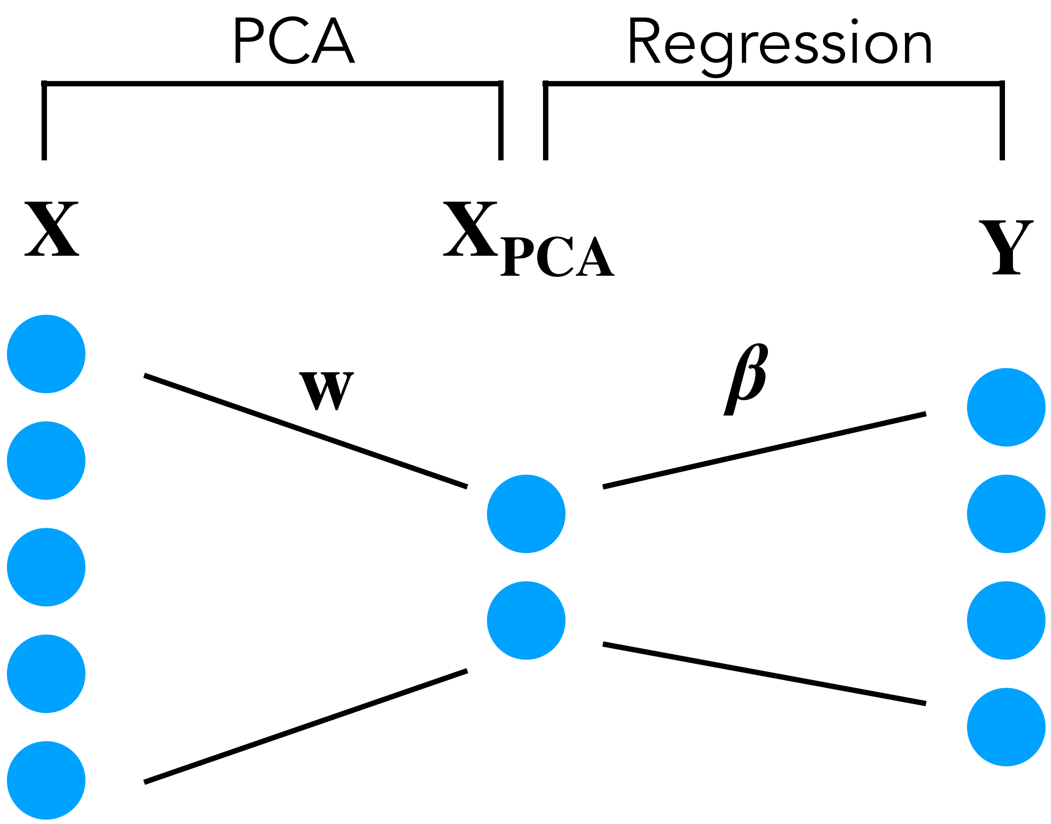
- PCA + Regression



- Dimensionality reduction does not take **Y** into account.

Principal Components Regression

- PCA + Regression



- Dimensionality reduction does not take **Y** into account.
- Dim. reduction acts like denoising / regularization

Reduced Rank Regression

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|^{2} \right)$$

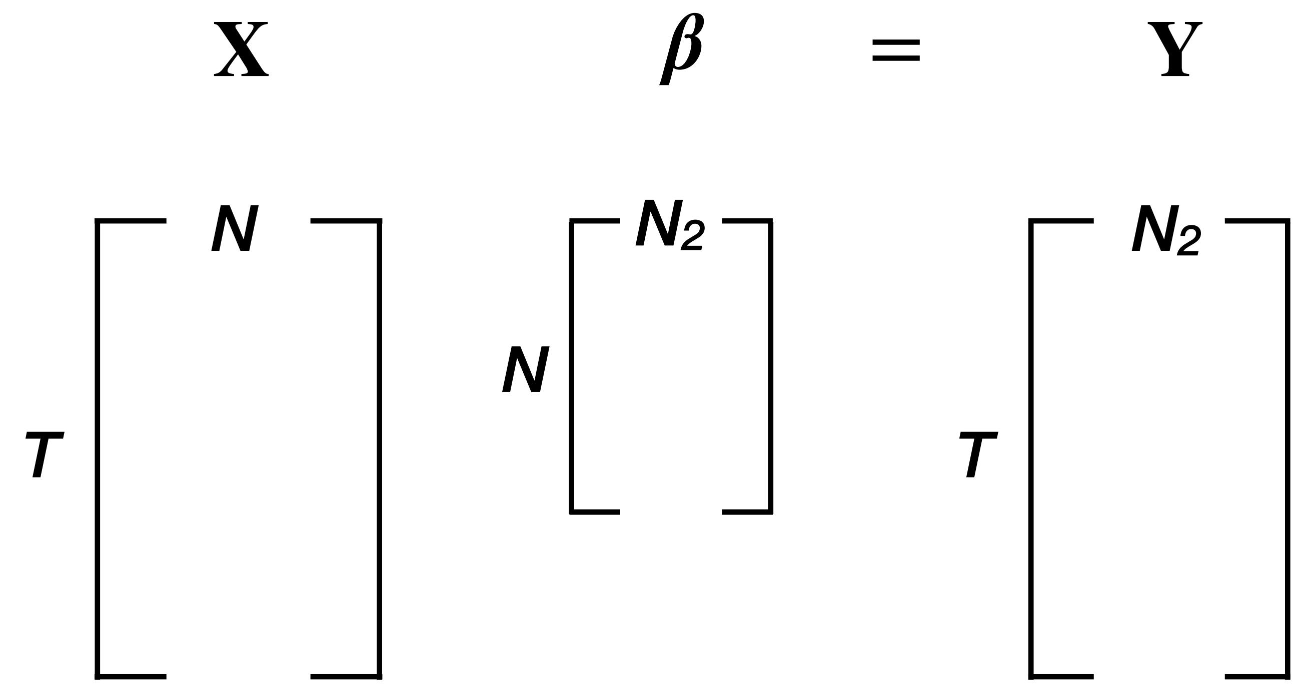
$$\mathbf{X} \qquad \boldsymbol{\beta} \qquad = \qquad \mathbf{Y}$$

$$\begin{array}{cc} & N \\ T & \left[\begin{array}{c} \quad \\ \quad \\ \quad \end{array}\right] \end{array} \quad \begin{array}{cc} & N_2 \\ N & \left[\begin{array}{c} \quad \\ \quad \end{array}\right] \end{array} \quad \begin{array}{cc} & N_2 \\ T & \left[\begin{array}{c} \quad \\ \quad \\ \quad \end{array}\right] \end{array}$$

Reduced Rank Regression

$$\underset{\boldsymbol{\beta}\in\,\mathbb{R}^{N\times N2}}{\arg\min}\left(\,\|\,\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\,\|^{\,2}\right)$$

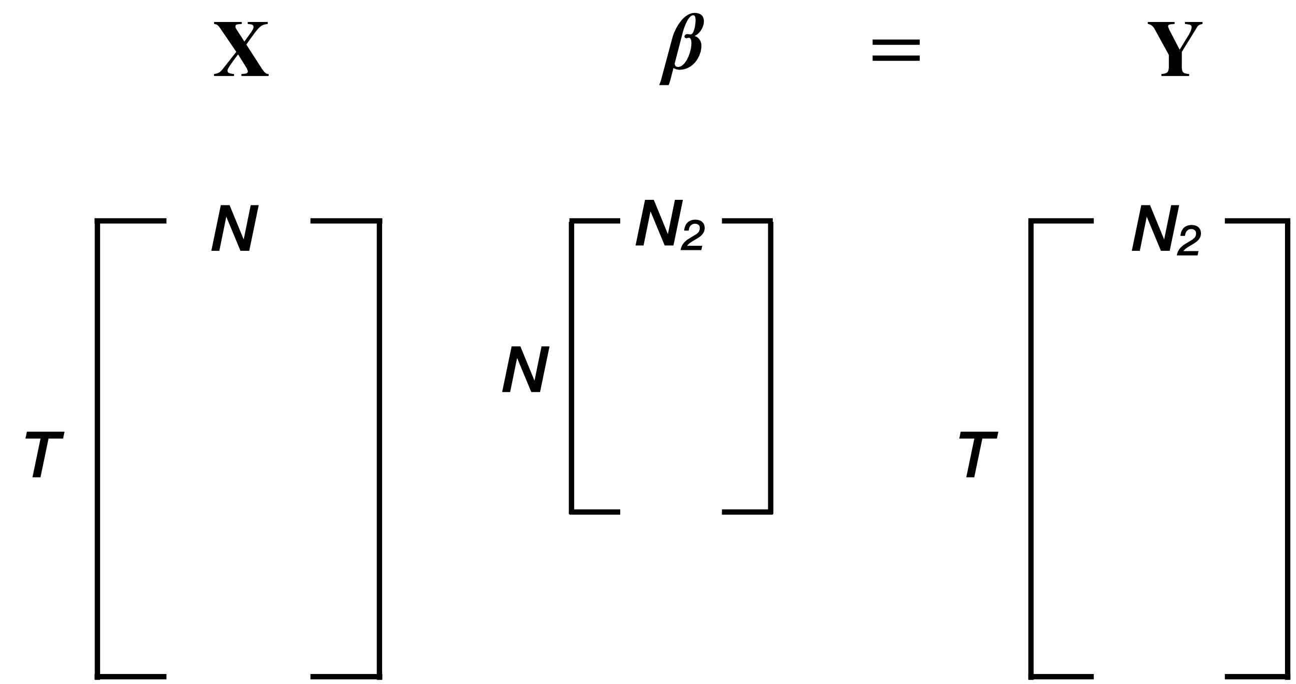
$$\textbf{rank}\left(\boldsymbol{\beta}\right)\leq K$$

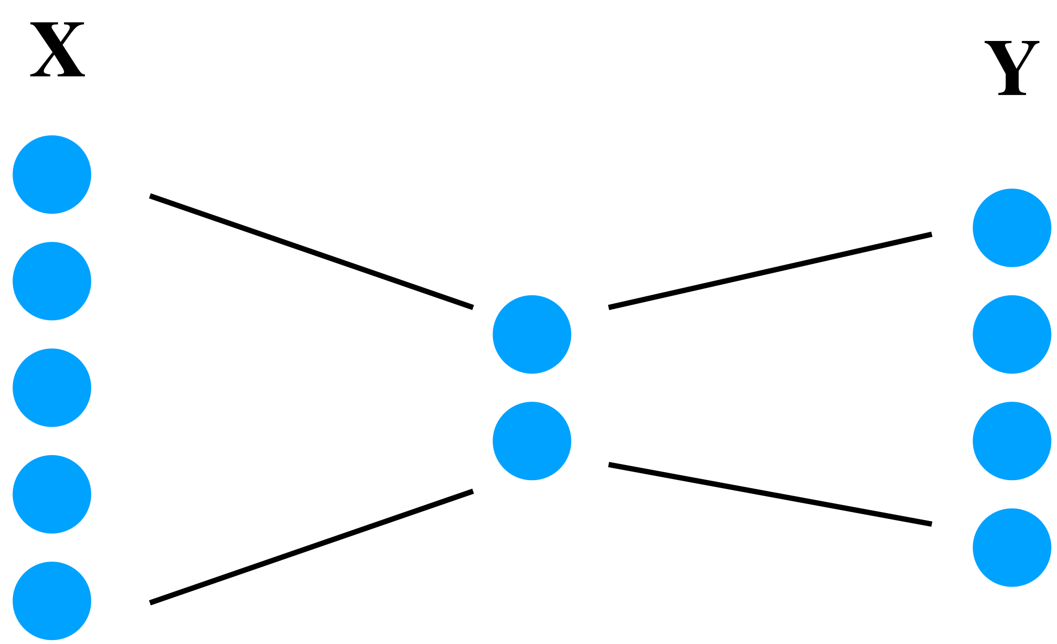


Reduced Rank Regression

$$\underset{\beta\in\,\mathbb{R}^{N\times N2}}{\arg\min}\left(\|\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|^{2}\right)$$

$$\textbf{rank}\left(\boldsymbol{\beta}\right)\leq K$$

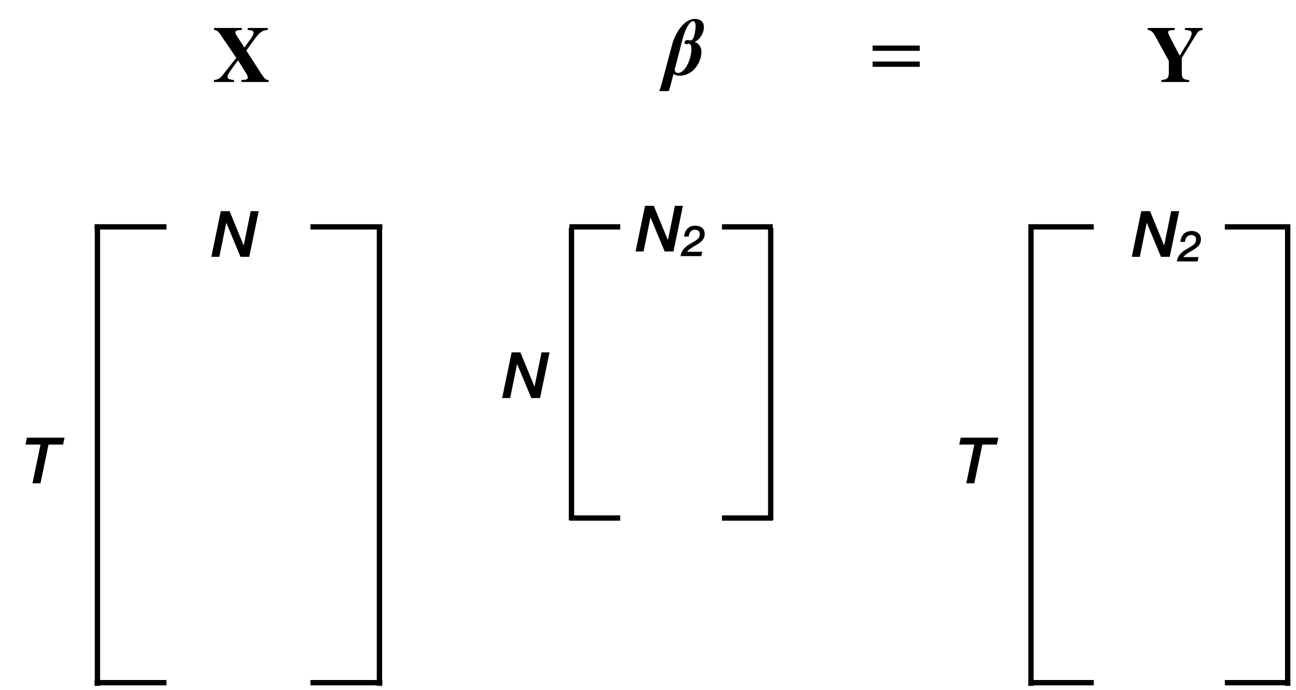




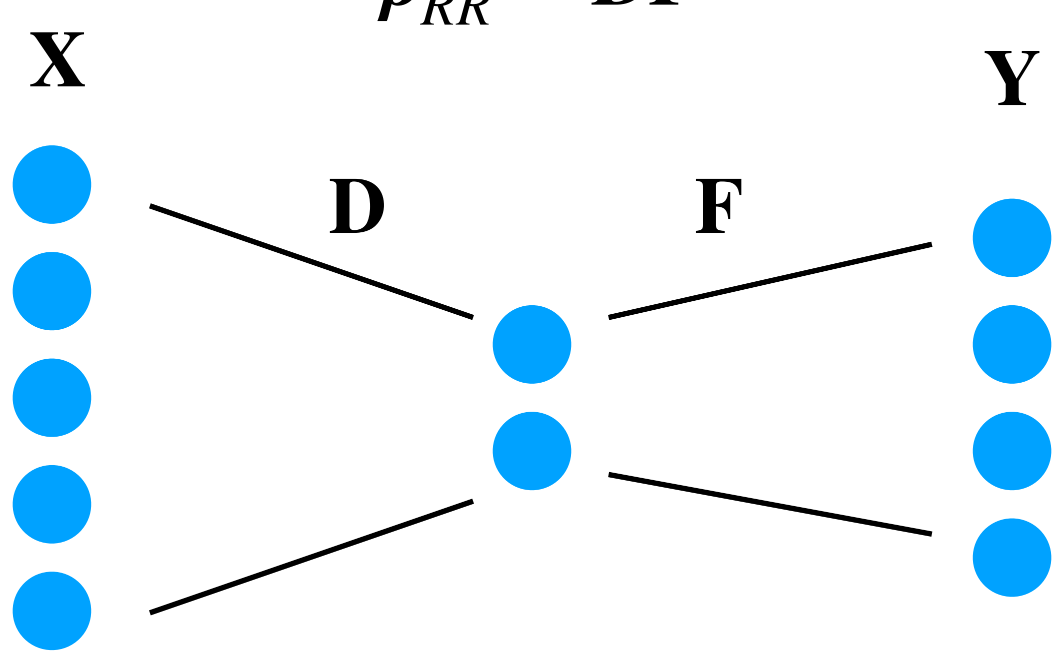
Reduced Rank Regression

$$\underset{\beta \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\beta \|^2 \right)$$

$$\textbf{rank}\left(\beta\right) \leq K$$



$$\beta_{RR} = \mathbf{DF}$$



Reduced Rank Regression Solution

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

Reduced Rank Regression Solution

$$\underset{\beta_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\right) + \left(\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}}\right) \|^2 =$$

Reduced Rank Regression Solution

$$\underset{\beta_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 =$$

$$\left\| \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\right) + \left(\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}}\right) \right\|^2 =$$

$$\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \right\|^2 + \left\| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 + \textbf{crossterms}$$

Reduced Rank Regression Solution

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 =$$

$$\left\| \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\right) + \left(\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}}\right) \right\|^2 =$$

$$\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \right\|^2 + \left\| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 + \cancelto{0}{\textbf{crossterms}}$$

Because residuals $\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\right)$ are orthogonal to **X**

Reduced Rank Regression Solution

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \|^2 + \| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2$$

Reduced Rank Regression Solution

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \le K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \|^2 + \| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2$$

Least squares loss

Loss from
low-rank approximation

Reduced Rank Regression Solution

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \|^2 + \| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2$$

Not a function of $\boldsymbol{\beta}_{RR}$

Reduced Rank Regression Solution

$$\underset{\beta_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\,(\boldsymbol{\beta}_{RR}) \leq K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \|^2 + \| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2$$

Not a function of $\boldsymbol{\beta}_{RR}$

Best rank K approximation is $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$

where **U** are the first K singular vectors of $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}$ (Eckart-Young theorem)

Reduced Rank Regression Solution

$$\underset{\beta_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\,(\boldsymbol{\beta}_{RR}) \leq K$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 =$$

$$\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} \|^2 + \| \mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2$$

$$\boldsymbol{\beta}_{\mathbf{RR}} = \boldsymbol{\beta}_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$$

Not a function of $\boldsymbol{\beta}_{RR}$

Best rank *K* approximation is $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$

where **U** are the first *K* singular vectors of $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}$ (Eckart-Young theorem)

Reduced Rank Regression Additional Notes

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$$\boldsymbol{\beta}_{\mathbf{RR}} = \boldsymbol{\beta}_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$$ where **U** are the first *K* singular vectors of $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}$

$$= \mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

- Dimensionality reduction depends on **Y**, unlike principal components regression

Reduced Rank Regression Additional Notes

$$\underset{\boldsymbol{\beta}_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\left\| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_{\mathbf{RR}} \right\|^2 \right) \qquad \textbf{rank}\left(\boldsymbol{\beta}_{RR}\right) \leq K$$

$\boldsymbol{\beta}_{\mathbf{RR}} = \boldsymbol{\beta}_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$ where **U** are the first K singular vectors of $\mathbf{X}\boldsymbol{\beta}_{\mathbf{OLS}}$

$$= \mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

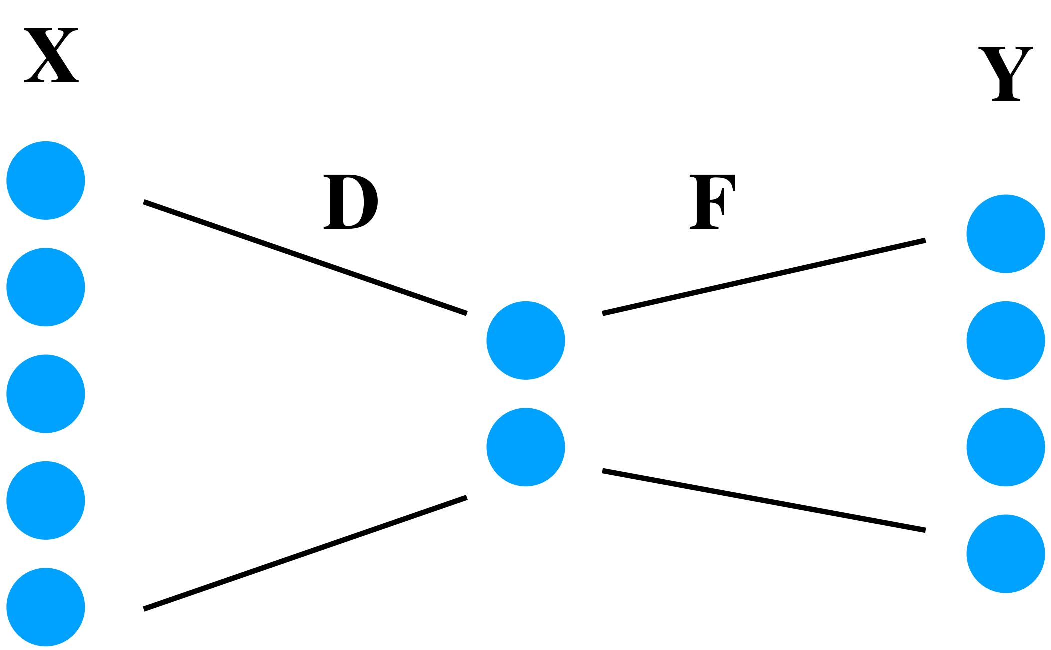
- Dimensionality reduction depends on **Y**, unlike principal components regression
- Reduced Rank Regression turns into PCA when **Y** = **X**.

Reduced Rank Regression Fun-facts

$$\underset{\beta_{RR} \in \mathbb{R}^{N \times N2}}{\arg\min} \left(\| \mathbf{Y} - \mathbf{X}\beta_{\mathbf{RR}} \|^2 \right) \qquad \textbf{rank}\left(\beta_{RR}\right) \leq K$$

$\beta_{\mathbf{RR}} = \beta_{\mathbf{OLS}}\mathbf{U}\mathbf{U}^{\mathbf{T}}$ where **U** are the first *K* singular vectors of $\mathbf{X}\beta_{\mathbf{OLS}}$

$$= \mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$



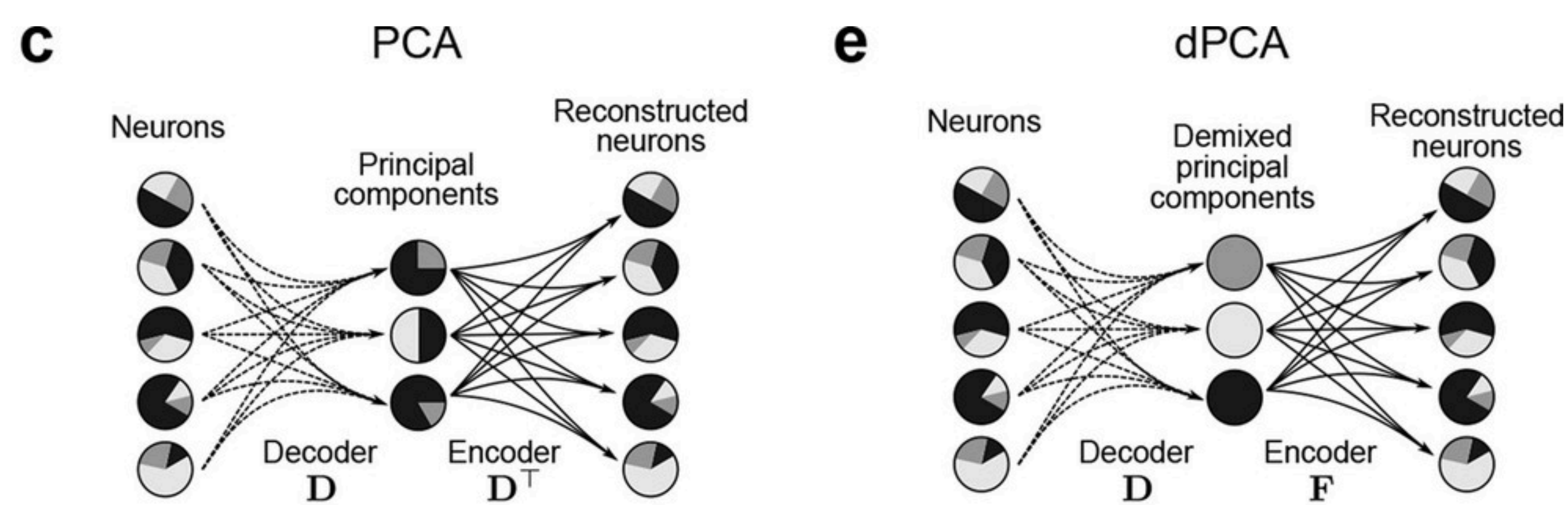
$$\beta_{RR} = \mathbf{DF}$$

$$\mathbf{D} = \beta_{\mathbf{OLS}}\mathbf{U}$$

$$\mathbf{F} = \mathbf{U}^{\mathbf{T}}$$

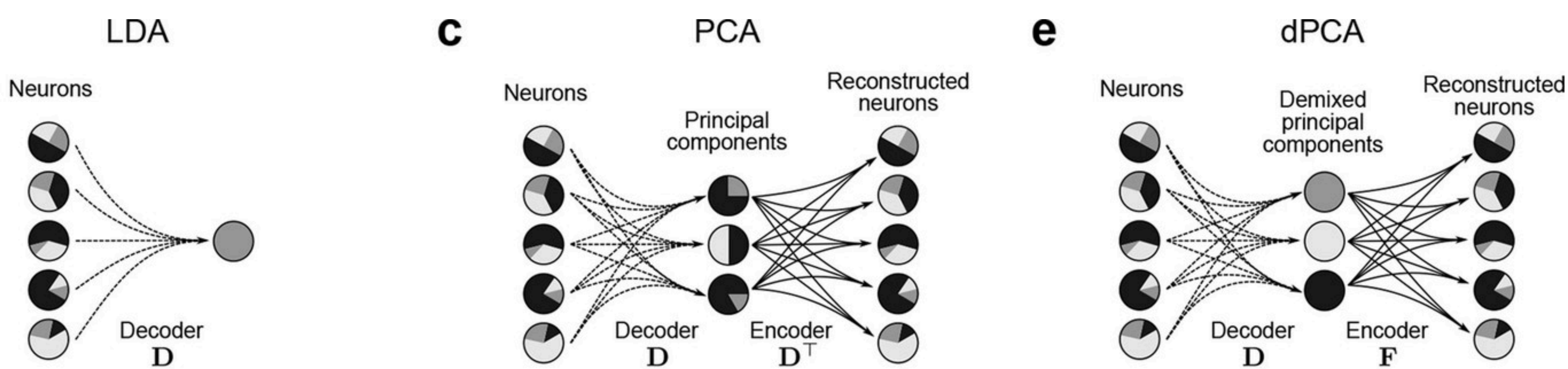
Demixed PCA

- Aims to find a latent representation that both reconstructs neural data (as in PCA) and maps the latents onto task parameters.
- Kobak,…, Machens. eLife (2016).

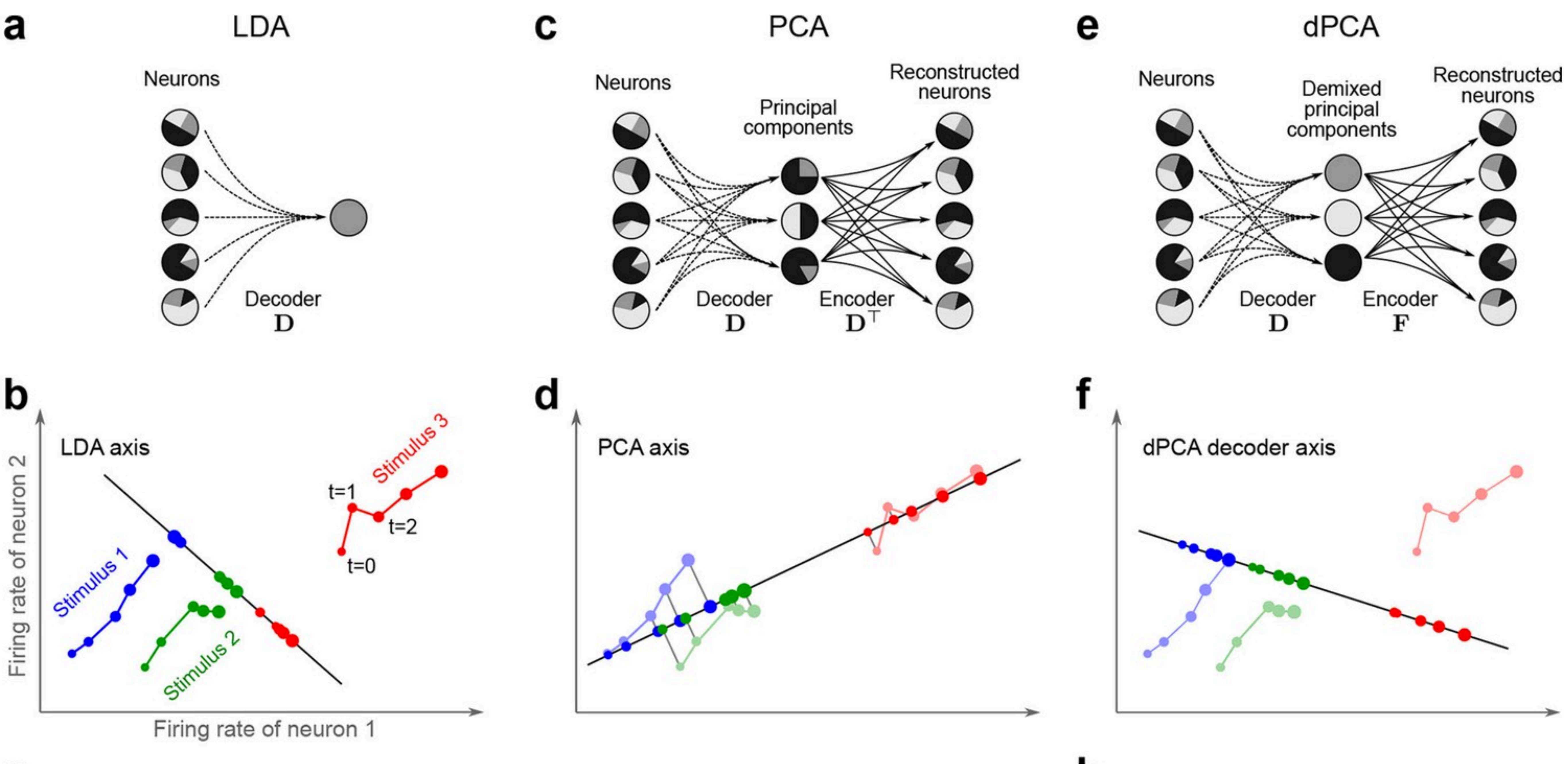


Demixed PCA

- Aims to find a latent representation that both reconstructs neural data (as in PCA) and maps the latents onto task parameters.
- Kobak,…, Machens. eLife (2016).

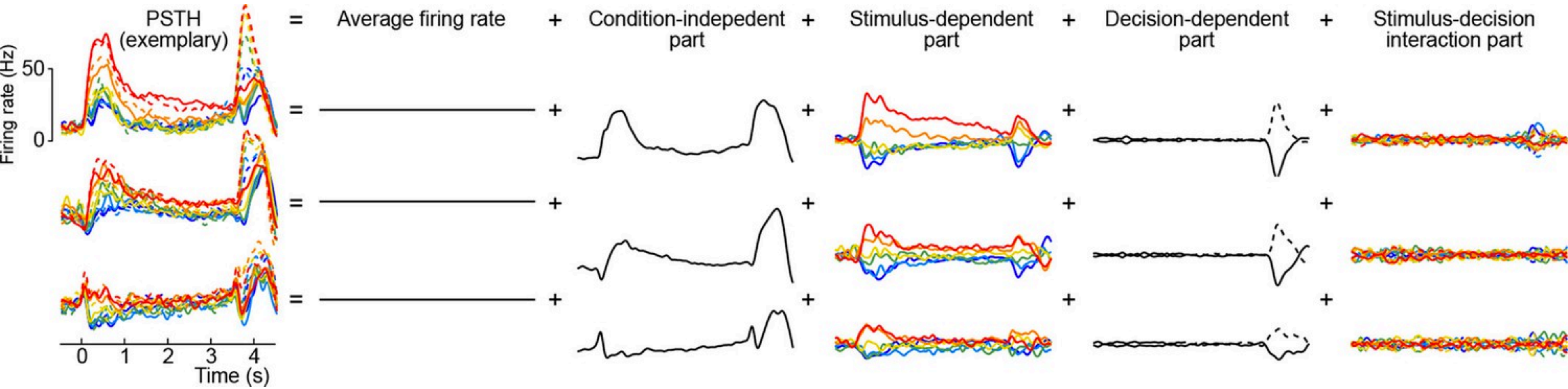


Demixed PCA



a
LDA
Neurons
Decoder
D
c
PCA
Neurons
Principal components
Reconstructed neurons
Decoder
D
Encoder
D^⊤
e
dPCA
Neurons
Demixed principal components
Reconstructed neurons
Decoder
D
Encoder
F
b
LDA axis
Stimulus 1
Stimulus 2
Stimulus 3
t=0
t=1
t=2
Firing rate of neuron 2
Firing rate of neuron 1
d
PCA axis
f
dPCA decoder axis

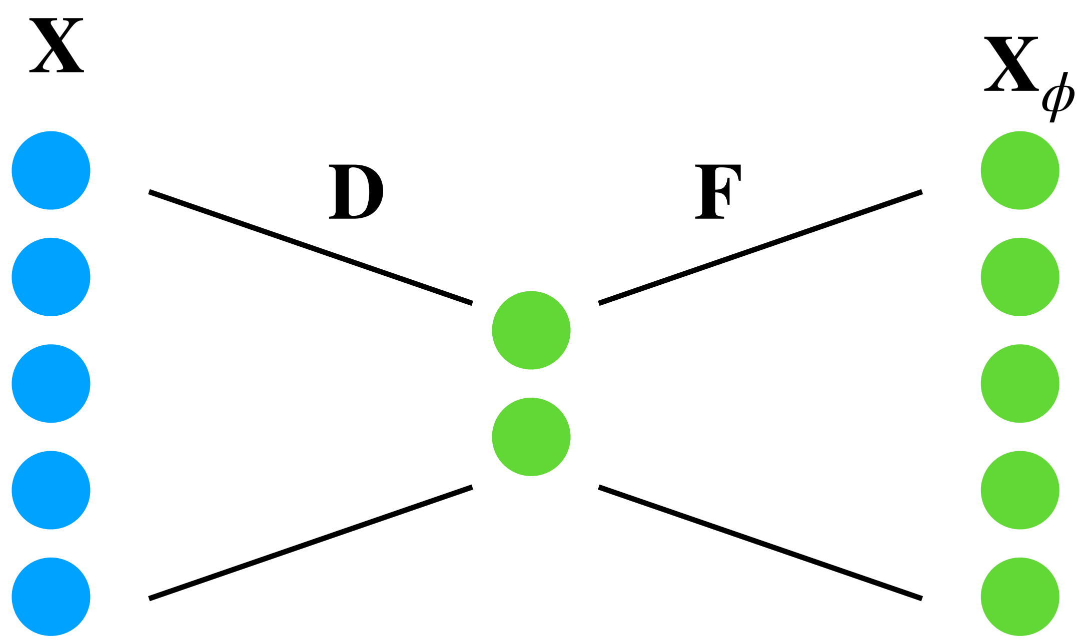
Demixed PCA



$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

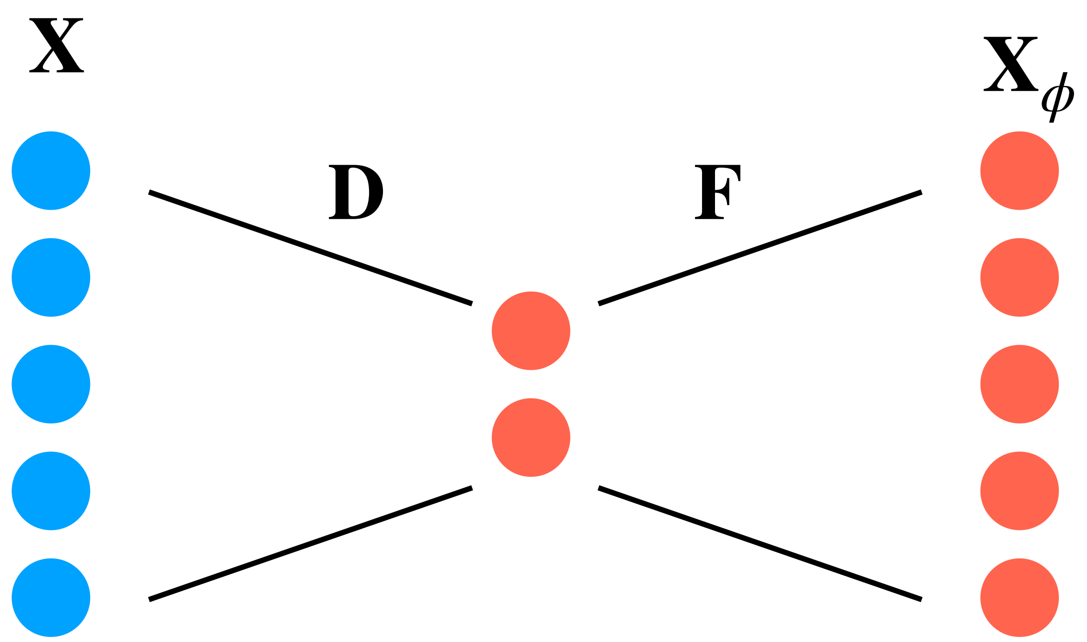
$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi} \mathbf{D}_{\phi} \mathbf{X} \right\|^2.$$



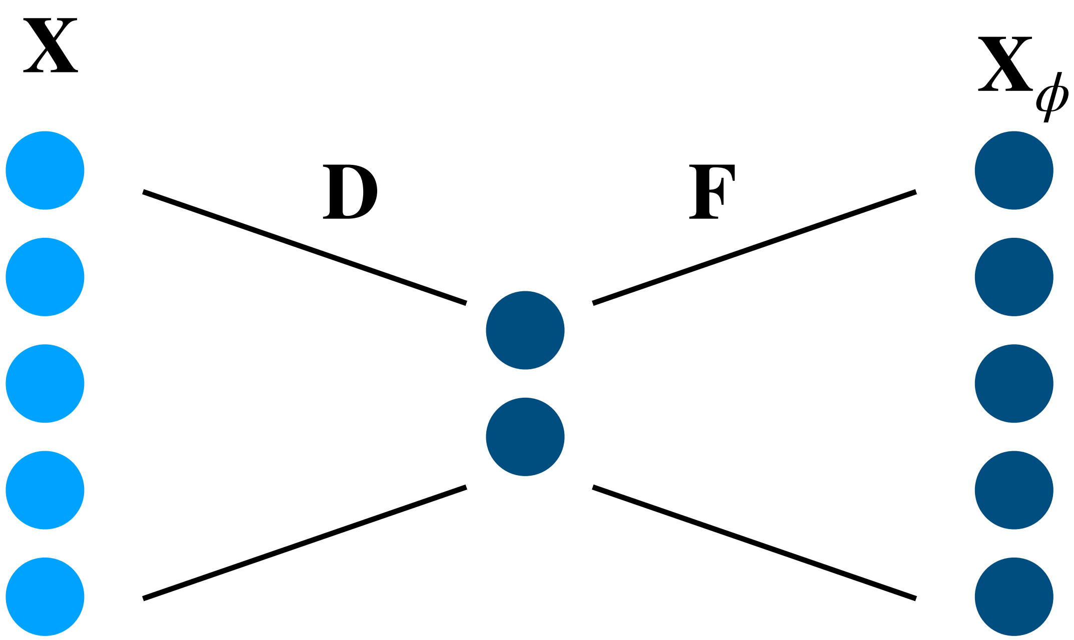
$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi} \mathbf{D}_{\phi} \mathbf{X} \right\|^2.$$



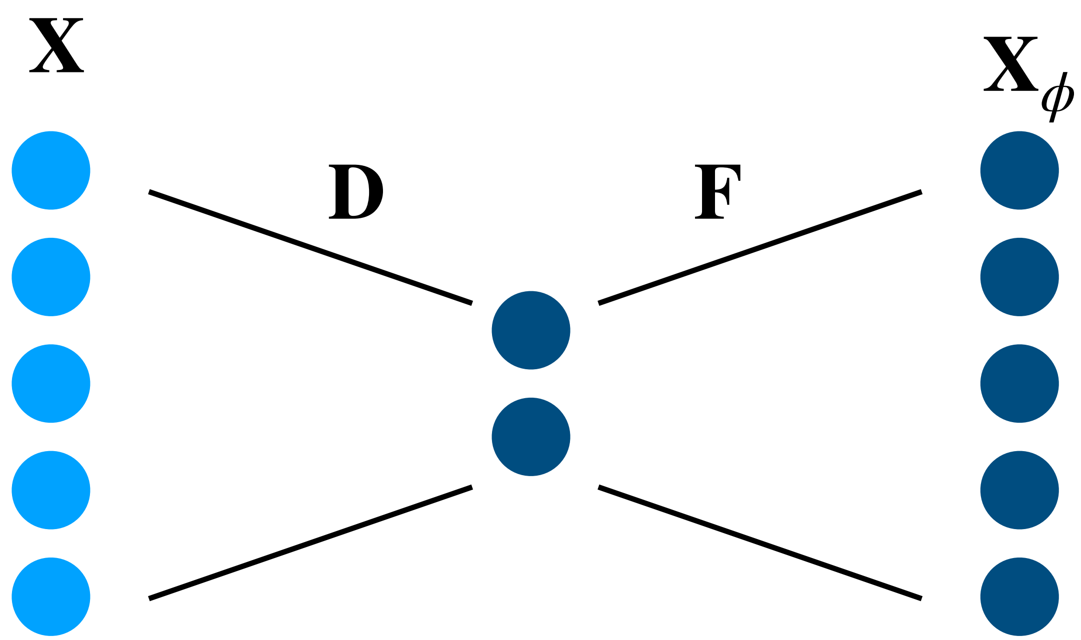
$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi} \mathbf{D}_{\phi} \mathbf{X} \right\|^2.$$



$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi}\mathbf{D}_{\phi}\mathbf{X} \right\|^2.$$



$$\beta_{RR} = \mathbf{DF}$$

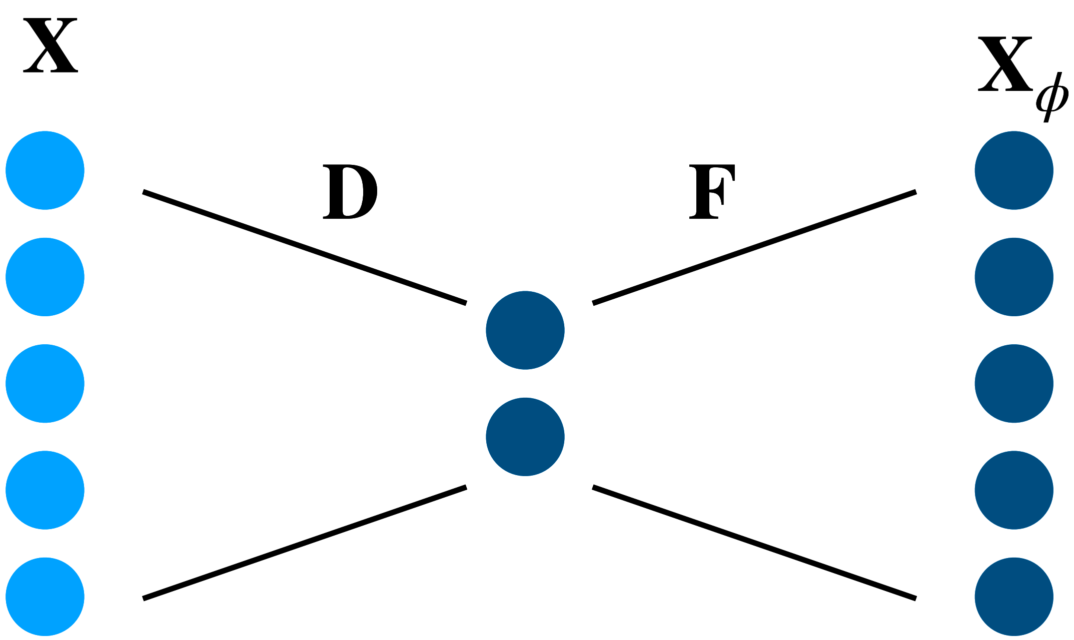
$$\mathbf{D} = \beta_{\mathbf{OLS}}\mathbf{U}$$

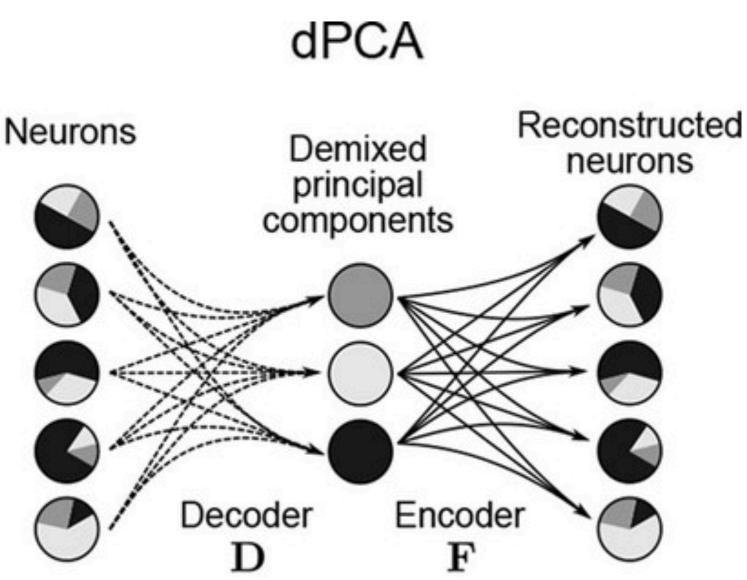
$$\mathbf{F} = \mathbf{U}^{\mathbf{T}}$$

- Just do Reduced Rank Regression for each condition!

$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

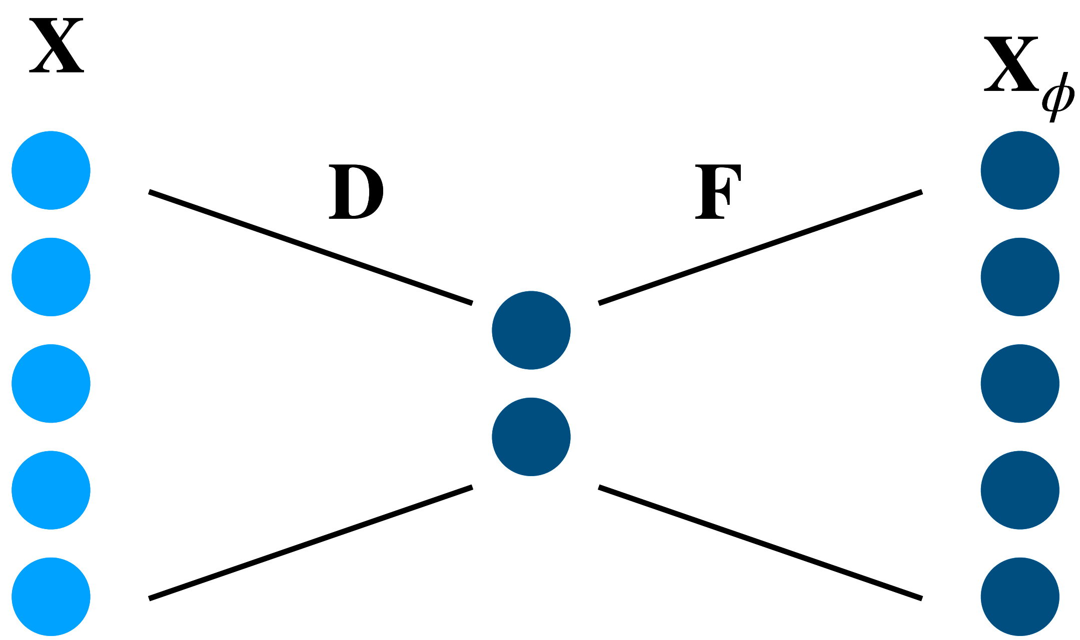
$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi}\mathbf{D}_{\phi}\mathbf{X} \right\|^2.$$

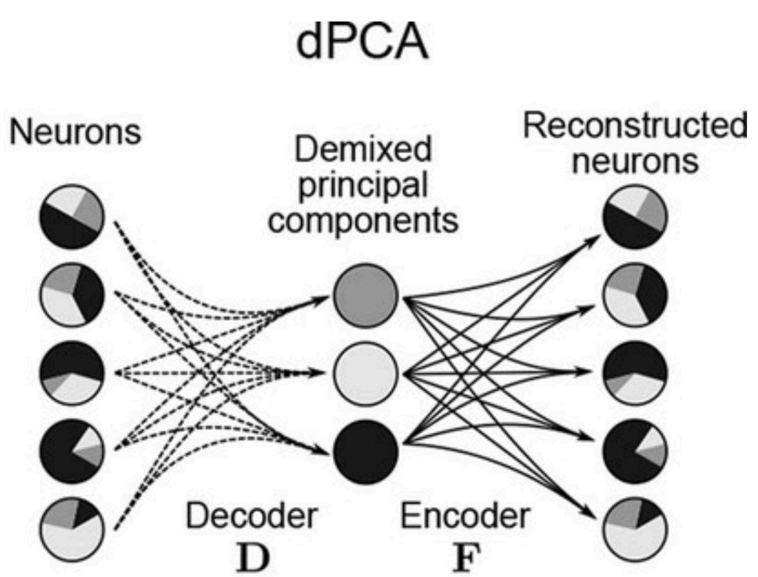




$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_{st} + \mathbf{X}_{dt} + \mathbf{X}_{sdt} + \mathbf{X}_{\text{noise}} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}.$$

$$L_{\text{dPCA}} = \sum_{\phi} \left\| \mathbf{X}_{\phi} - \mathbf{F}_{\phi} \mathbf{D}_{\phi} \mathbf{X} \right\|^2.$$



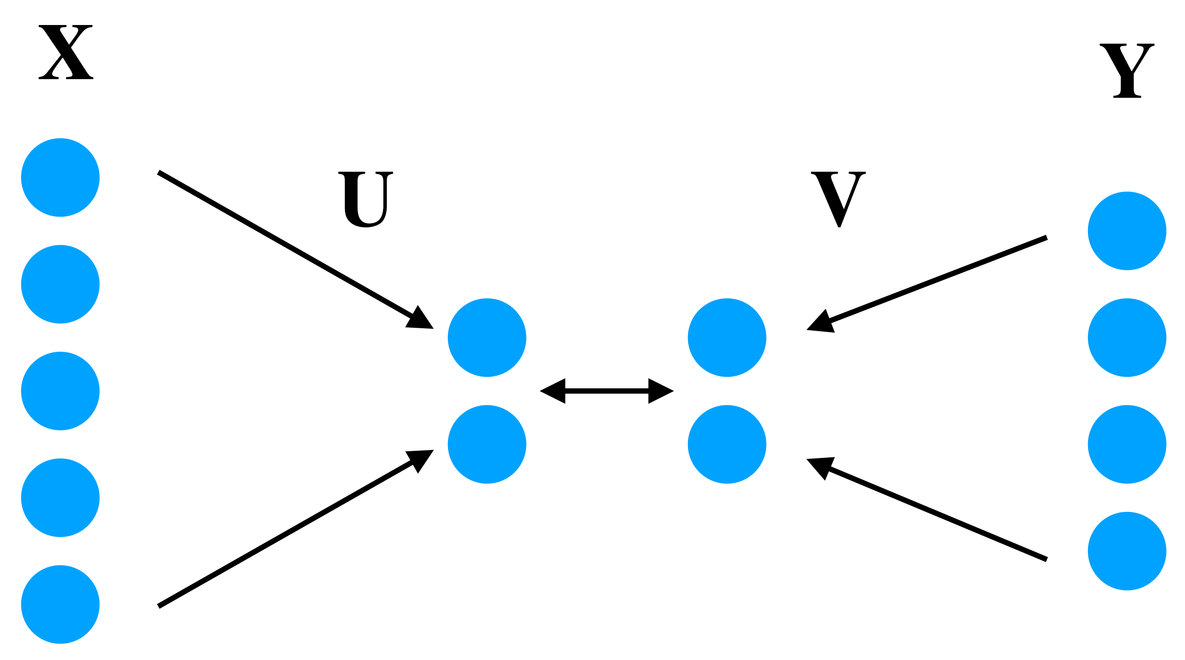


- Related methods include:
  - Targeted dimensionality reduction (Mante,…,Newsome, 2013)
  - Model-based targeted dimensionality reduction (Aoi,…,Pillow, 2019)

Overview

1. Reducing dimensionality in one space
   - PCA and regression refreshers
   - Principal components regression
   - Reduced rank regression
   - Demixed PCA
2. Reducing dimensionality in two spaces
   - Canonical correlation analysis
   - Partial least squares

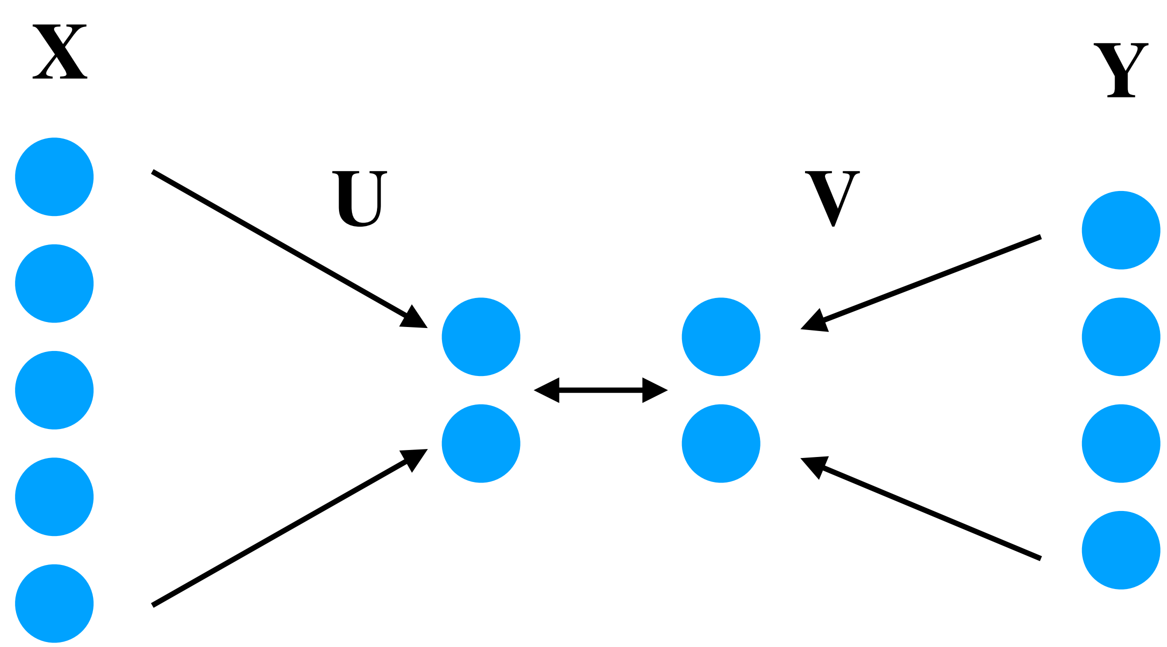
Reducing dimensionality in two spaces



Partial Least Squares

- Maximize covariance between $\mathbf{XU}$ and $\mathbf{YV}$

$$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^2 \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \ \| \mathbf{V} \| = \mathbf{I}$$

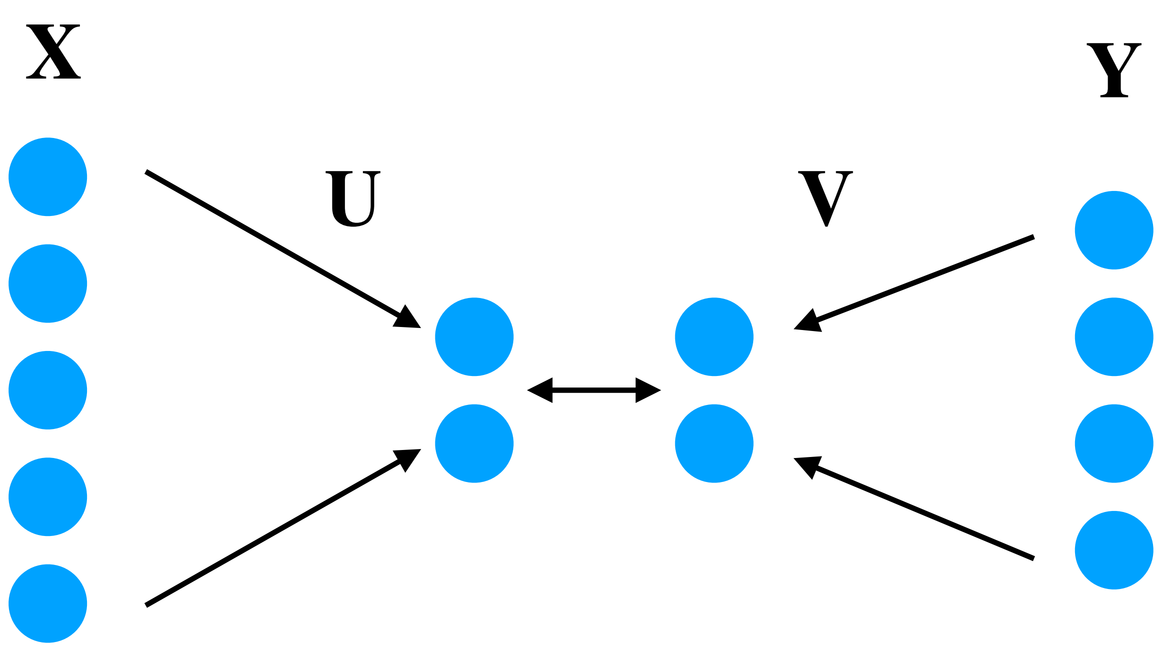


Partial Least Squares

- Maximize covariance between $\mathbf{XU}$ and $\mathbf{YV}$

$$\arg\max_{\mathbf{U},\mathbf{V}} \left(\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^2 \right) \textbf{ s.t. } \|\mathbf{U}\| = \mathbf{I}, \|\mathbf{V}\| = \mathbf{I}$$

- Solution is to take first *K* singular vectors of $\mathbf{X}^{\mathbf{T}}\mathbf{Y}$
  - **U** = left singular vectors, **V** = right singular vectors



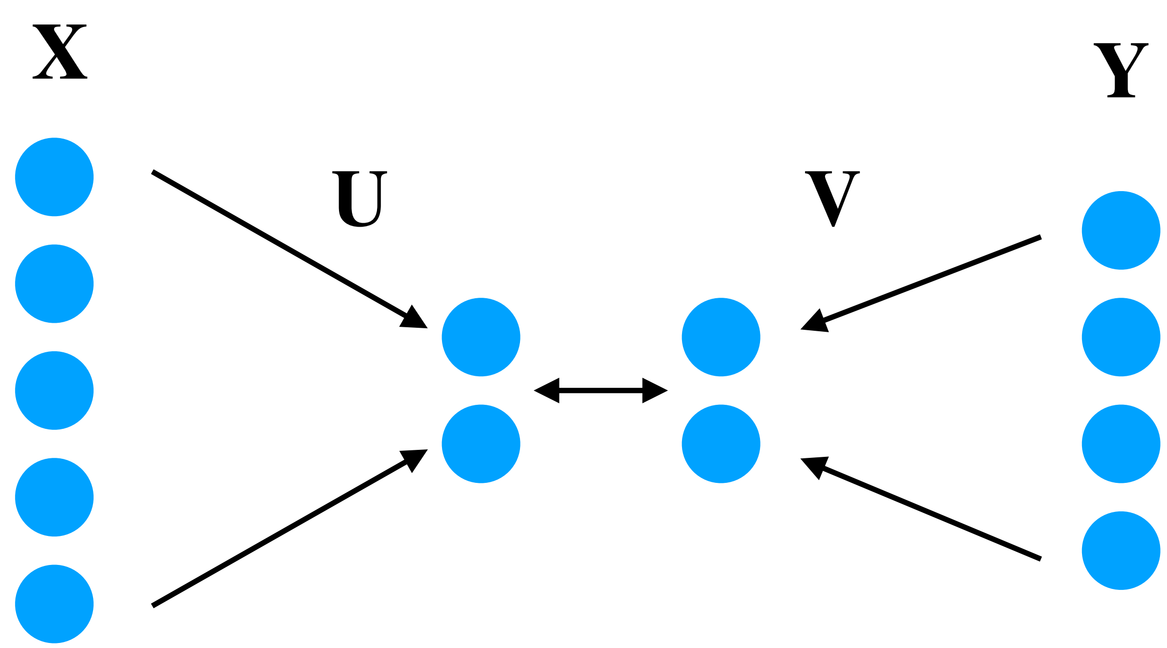
Partial Least Squares

- Maximize covariance between $\mathbf{XU}$ and $\mathbf{YV}$

$$\arg\max_{\mathbf{U},\mathbf{V}} \left(\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^2 \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \| \mathbf{V} \| = \mathbf{I}$$

- Minimize error between $\mathbf{XU}$ and $\mathbf{YV}$

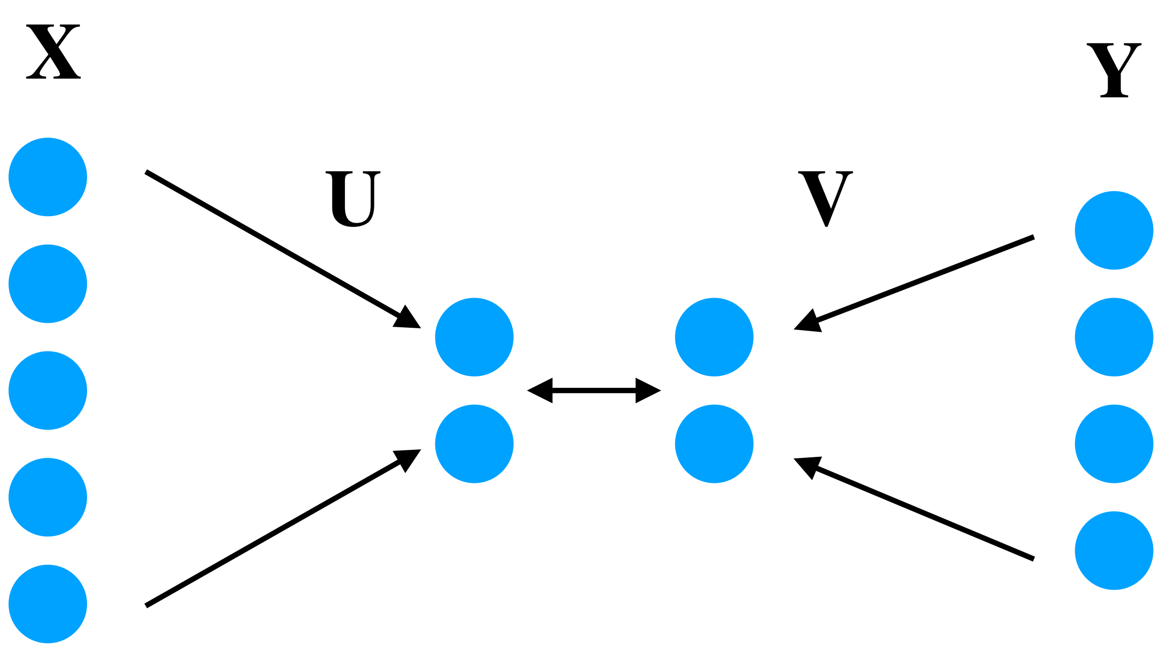
$$\arg\min_{\mathbf{U},\mathbf{V}} \left(\| \mathbf{XU} - \mathbf{YV} \|^2 \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \| \mathbf{V} \| = \mathbf{I}$$



Canonical Correlation Analysis

- Maximize correlation between $\mathbf{XU}$ and $\mathbf{YV}$

$$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\frac{\|(\mathbf{XU})^{\mathbf{T}}\mathbf{YV}\|^{2}}{\|\mathbf{XU}\|\ \|\mathbf{YV}\|}\right)\ \textbf{s.t.}\ \|\mathbf{U}\| = \mathbf{I},\ \|\mathbf{V}\| = \mathbf{I}$$

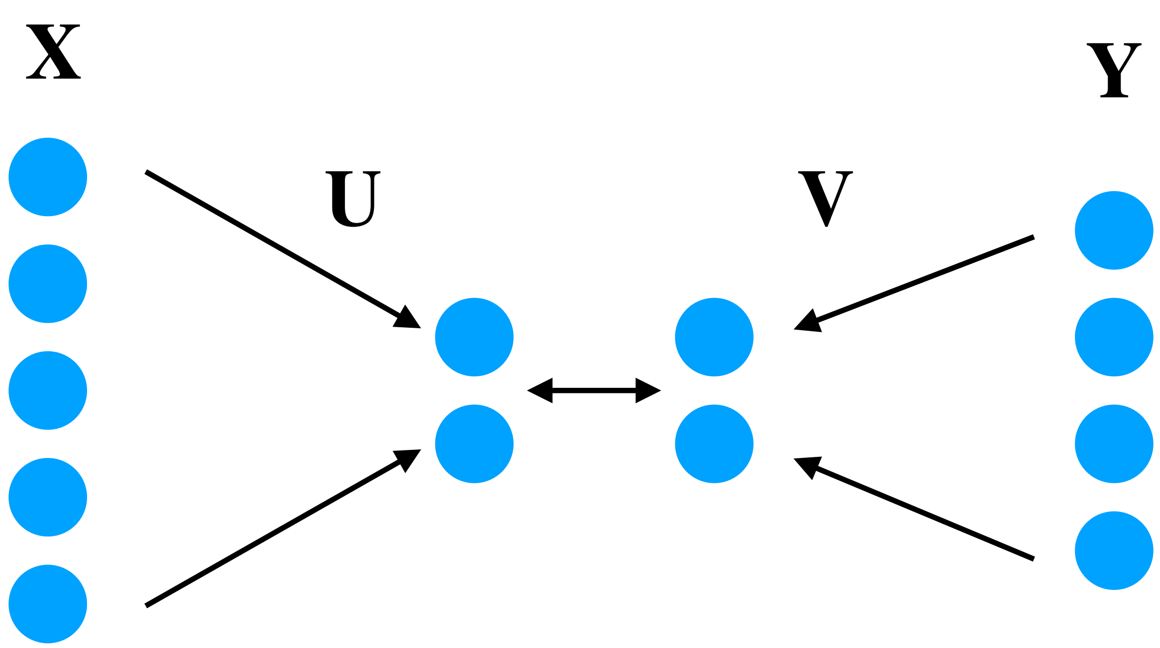


Canonical Correlation Analysis

- Maximize correlation between $\mathbf{XU}$ and $\mathbf{YV}$

$$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\frac{\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^{2}}{\| \mathbf{XU} \| \; \| \mathbf{YV} \|} \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \; \| \mathbf{V} \| = \mathbf{I}$$

- Solution is to take first *K* singular vectors of $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Y}(\mathbf{Y}^{\mathbf{T}}\mathbf{Y})^{-1/2}$
  - **U** = left singular vectors, **V** = right singular vectors



CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\left\|(\mathbf{X}\mathbf{U})^{\mathbf{T}}\mathbf{Y}\mathbf{V}\right\|^{2}\right) \text{ s.t. } \|\mathbf{U}\| = \mathbf{I},\ \|\mathbf{V}\| = \mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\frac{\left\|(\mathbf{X}\mathbf{U})^{\mathbf{T}}\mathbf{Y}\mathbf{V}\right\|^{2}}{\|\mathbf{X}\mathbf{U}\|\ \|\mathbf{Y}\mathbf{V}\|}\right) \text{ s.t. } \|\mathbf{U}\| = \mathbf{I},\ \|\mathbf{V}\| = \mathbf{I}$$

CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\left\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \right\|^{2} \right) \textbf{ s.t. } \left\| \mathbf{U} \right\| = \mathbf{I},\ \left\| \mathbf{V} \right\| = \mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\left\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \right\|^{2} \right) \textbf{ s.t. } \left\| \mathbf{XU} \right\| = \mathbf{I},\ \left\| \mathbf{YV} \right\| = \mathbf{I}$$

CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\min}\left(\|\mathbf{XU}-\mathbf{YV}\|^{2}\right)\ \textbf{s.t.}\ \|\mathbf{U}\|=\mathbf{I},\ \|\mathbf{V}\|=\mathbf{I}$$

CCA:

CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\min}\left(\|\mathbf{XU}-\mathbf{YV}\|^{2}\right)\ \textbf{s.t.}\ \|\mathbf{U}\|=\mathbf{I},\ \|\mathbf{V}\|=\mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\min}\left(\|\mathbf{XU}-\mathbf{YV}\|^{2}\right)\ \textbf{s.t.}\ \|\mathbf{XU}\|=\mathbf{I},\ \|\mathbf{YV}\|=\mathbf{I}$$

CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^{2} \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \| \mathbf{V} \| = \mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\frac{\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \|^{2}}{\| \mathbf{XU} \| \; \| \mathbf{YV} \|} \right) \textbf{ s.t. } \| \mathbf{U} \| = \mathbf{I}, \| \mathbf{V} \| = \mathbf{I}$$

PLS: Solution is to take first *K* singular vectors of $\mathbf{X}^{\mathbf{T}}\mathbf{Y}$

CCA: Solution is to take first *K* singular vectors of $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Y}(\mathbf{Y}^{\mathbf{T}}\mathbf{Y})^{-1/2}$

CCA vs PLS

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\left\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \right\|^2 \right) \textbf{ s.t. } \left\| \mathbf{U} \right\| = \mathbf{I}, \left\| \mathbf{V} \right\| = \mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max} \left(\frac{\left\| (\mathbf{XU})^{\mathbf{T}}\mathbf{YV} \right\|^2}{\left\| \mathbf{XU} \right\| \left\| \mathbf{YV} \right\|} \right) \textbf{ s.t. } \left\| \mathbf{U} \right\| = \mathbf{I}, \left\| \mathbf{V} \right\| = \mathbf{I}$$

PLS: Solution is to take first *K* singular vectors of $\mathbf{X}^{\mathbf{T}}\mathbf{Y}$

CCA: Solution is to take first *K* singular vectors of $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Y}(\mathbf{Y}^{\mathbf{T}}\mathbf{Y})^{-1/2}$

- CCA solution “whitens” **X** and **Y**

CCA vs PLS vs RRR

PLS: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\left\|(\mathbf{XU})^{\mathbf{T}}\mathbf{YV}\right\|^2\right)\ \textbf{s.t.}\ \|\mathbf{U}\| = \mathbf{I},\ \|\mathbf{V}\| = \mathbf{I}$$

CCA: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\frac{\left\|(\mathbf{XU})^{\mathbf{T}}\mathbf{YV}\right\|^2}{\|\mathbf{XU}\|\ \ \|\mathbf{YV}\|}\right)\ \textbf{s.t.}\ \|\mathbf{U}\| = \mathbf{I},\ \|\mathbf{V}\| = \mathbf{I}$$

RRR: $$\underset{\mathbf{U},\mathbf{V}}{\arg\max}\left(\frac{\left\|(\mathbf{XU})^{\mathbf{T}}\mathbf{YV}\right\|^2}{\|\mathbf{XU}\|}\right)\ \textbf{s.t.}\ \|\mathbf{V}\| = \mathbf{I}$$

- CCA solution "whitens" **X** and **Y**
- RRR solution "whitens" **X**, but not **Y**

$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

Many extensions

- Adding regularization
- Reduced Rank GLMs
- Kernel CCA
- Neural networks with bottlenecks
- Multiple factor analysis (for >2 populations)
- Probabilistic versions (pCCA)