Overview

- **Ordinary Least Squares (OLS) Regression**: Finds the projection direction of x's that are maximally correlated with the y's
- **PCA:** Finds projection directions of the x's with maximal covariance
- **Principal Component Regression (PCR):** Does PCA for dimensionality reduction on X, and then OLS using PC features**.**
- **Canonical Covariance Analysis:** Finds the projection directions of X and Y that maximize their mutual *covariance*.
- **Canonical Correlation Analysis (CCA):** Finds the projection directions of X and Y that maximize their mutual *correlation*.
- **All use SVD to minimize reconstruction error or maximize variance/covariance**

PCA

- $XV = Z$
- $X = ZV^{+} = ZV^{T}$

Z

$X = ZV^{T}$ ↓ ↑ $Z = XV$

X

Canonical Covariance Analysis

If Y is high-dimensional, we might want to do dimension reduction for *both* Y and X.

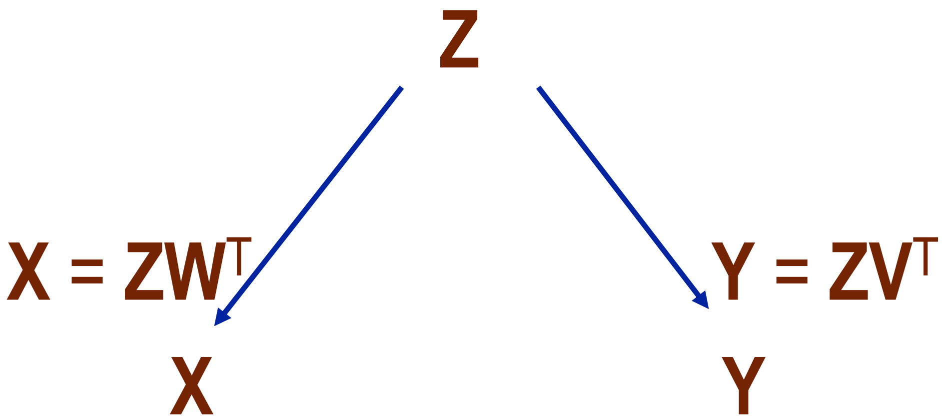
C*anonical covariance analysis* finds the projection directions for both X and Y to maximize their covariance.

or to best reconstruct **X** from **Y** and to reconstruct **Y** from **X**

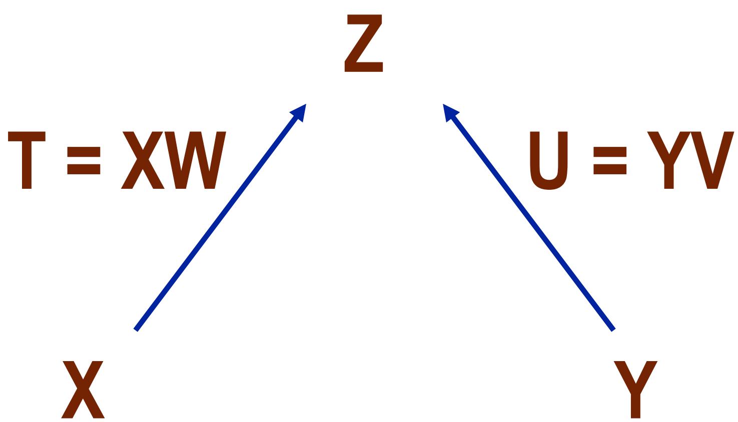
(Comparison: PCA finds the projection directions of maximum covariance for X with itself.)

This is one type of Partial Least Squares (PLS), which find projections of x that explain all the y's.

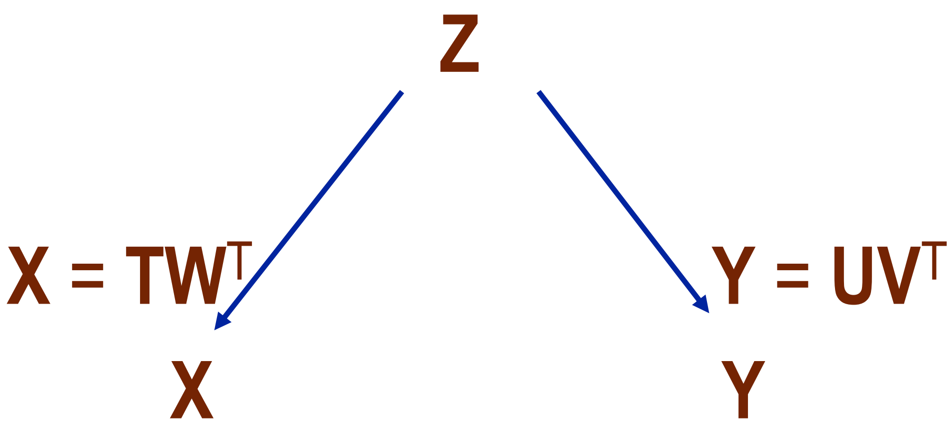
Canonical Covariance Analysis



Canonical Covariance Analysis



Canonical Covariance Analysis



Canonical Covariance Analysis

Find reduced-dimension representations **T** (of $\mathbf{X}_c$) and **U** (of $\mathbf{Y}_c$) such that each pair of corresponding columns $\mathbf{t}_i$ **,** $\mathbf{u}_i$ are optimal in the following sense:

Let $$w_i^*,\ v_i^* = \underset{\substack{w_i \in \mathbb{R}^p,\, v_i \in \mathbb{R}^m \\ w_i^T w_i = v_i^T v_i = 1}}{argmax} \ (X_c w_i)^T (Y_c v_i)$$

Subject to: $(X_c w_i^*)^T (X_c w_j^*) = 0$ for all **j**<**i**.

Then: $t_i := X_c w_i^*$ and $u_i := Y_c v_i^*$

What are the scores? The loadings?

Canonical Covariance Analysis

$$\operatorname{argmax} \quad (X_c w_i)^T (Y_c v_i)$$

- $(Xw)^T(Yv)$
- $w^T(X^TY)v$

Finds the left and right singular vectors of X^TY

Canonical Correlation Analysis (CCA)

Uses the singular vectors of: $(X^T X)^{-1/2} X^T Y (Y^T Y)^{-1/2}$

Correlation: re-scales the data, no units. Range -1 to 1.

Analog to auto-scaling: if $\mathbf{X}^T\mathbf{X}$ is diagonal, then this divides each row of $\mathbf{X}^T$ by the corresponding diagonal element of $(\mathbf{X}^T\mathbf{X})^{1/2}$.

In the general case where $\mathbf{X}^T\mathbf{X}$ is not diagonal: this normalizes $\mathbf{X}^T$ by "removing" covariance.

"Whitens" the data.

CCA advantages

- **CCA is scale invariant**
  - Since X and Y are whitened,
- **CCA is symmetric in X and Y**

Recap

- **OLS**: Finds the projection direction of **x** maximally correlated with y
- **PCA:** Finds projection directions of the **X** with maximal covariance
  - SVD of **X'X**
- **Canonical Covariance Analysis:** Finds the projection directions of **X** and **Y** that maximize their *covariance*
  - SVD of **Y'X**
- **Canonical Correlation Analysis (CCA):** Finds the projection directions of X and Y that maximize their *correlation*
  - SVD of $(X'X)^{-1/2}X'Y(Y'Y)^{-1/2}$
  - The whitening makes it scale invariant

All minimize reconstruction error and maximize variance/covariance