Canonical Correlation Analysis
Audio might have background sounds uncorrelated with video

Video might have lighting changes uncorrelated with audio

Redundant information between two views: the speech
Canonical Correlation Analysis
Canonical Correlation Analysis

- Age
- Gender
- Candies per week
- Favorite Cartoon
Data comes in pairs \((x_1, x'_1), \ldots, (x_n, x'_n)\) where \(x_t\)'s are \(d\) dimensional and \(x'_t\)'s are \(d'\) dimensional.
Data comes in pairs \((x_1, x'_1), \ldots, (x_n, x'_n)\) where \(x_t\)'s are \(d\) dimensional and \(x'_t\)'s are \(d'\) dimensional.

Goal: Compress say view one into \(y_1, \ldots, y_n\), that are \(K\) dimensional vectors.

- Retain information redundant between the two views.
- Eliminate “noise” specific to only one of the views.
Method A and Method B are both equally good feature extraction techniques
Example II: Combining Feature Extractions

- Method A and Method B are both equally good feature extraction techniques.

- Concatenating the two features blindly yields a large dimensional feature vector with redundancy.
Method A and Method B are both equally good feature extraction techniques.

Concatenating the two features blindly yields large dimensional feature vector with redundancy.

Applying techniques like CCA extracts the key information between the two methods.
Method A and Method B are both equally good feature extraction techniques.

Concatenating the two features blindly yields large dimensional feature vector with redundancy.

Applying techniques like CCA extracts the key information between the two methods.

Removes extra unwanted information.
How do we get the right direction? (single dimension $K = 1$)
How do we get the right direction? (single dimension K = 1)

Age
+ Gender
Candies per week
Favorite Cartoon
Which Direction to Pick?

View I

View II
Which Direction to Pick?

View I

View II
Which Direction to Pick?

View I

View II
Which Direction to Pick?

View I

View II
Which Direction to Pick?

PCA direction
Direction has large covariance
How do we pick the right direction to project to?
Pick the directions in each view so that resultant projections have high covariance.
Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively. We want these directions to maximize,

$$\frac{1}{n} \sum_{t=1}^{n} \left( y_t[1] - \frac{1}{n} \sum_{t=1}^{n} y_t[1] \right) \cdot \left( y'_t[1] - \frac{1}{n} \sum_{t=1}^{n} y'_t[1] \right)$$

where $y_t[1] = \mathbf{w}_1^\top \mathbf{x}_t$ and $y'_t[1] = \mathbf{v}_1^\top \mathbf{x}'_t$
This should work right?!?!
How do we get the right direction? (single dimension K =1)
How do we get the right direction? (single dimension $K = 1$)

- Age
- Gender
- Candies per week
- Favorite Cartoon
Why not Maximize Covariance
WHY not MAXIMIZE COVARIANCE
Why not Maximize Covariance

Relevant information
**Why not Maximize Covariance**

Relevant information

Say covariance in some coordinate just happens to be $> 0$

Scaling up this coordinate we can blow up covariance
Say \( w_1 \) and \( v_1 \) are the directions we choose to project in views 1 and 2 respectively, we want these directions to maximize,

\[
\frac{\frac{1}{n} \sum_{t=1}^{n} (y_t[1] - \frac{1}{n} \sum_{t=1}^{n} y_t[1]) \cdot (y_t'[1] - \frac{1}{n} \sum_{t=1}^{n} y_t'[1])}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t[1] - \frac{1}{n} \sum_{t=1}^{n} y_t[1])^2} \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t'[1] - \frac{1}{n} \sum_{t=1}^{n} y_t'[1])^2}}
\]
Basic Idea of CCA

- Normalize variance in chosen direction to be constant (say 1)
- Then maximize covariance
- This is same as maximizing “correlation coefficient”
**Covariance Vs Correlation**

- **Covariance**\((A, B) = \mathbb{E}[(A - \mathbb{E}[A]) \cdot (B - \mathbb{E}[B])]\)

  Depends on the scale of \(A\) and \(B\). If \(B\) is rescaled, covariance shifts.

- **Correlation**\((A, B) = \frac{\mathbb{E}[(A-\mathbb{E}[A]) \cdot (B-\mathbb{E}[B])]}{\sqrt{\text{Var}(A)} \sqrt{\text{Var}(B)}}\)

  Scale free.
Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{1}{n} \sum_{t=1}^{n} \left( y_{t[1]} - \frac{1}{n} \sum_{t=1}^{n} y_{t[1]} \right) \cdot \left( y'_{t[1]} - \frac{1}{n} \sum_{t=1}^{n} y'_{t[1]} \right)$$

where $y_{t[1]} = \mathbf{w}_1^\top \mathbf{x}_t$ and $y'_{t[1]} = \mathbf{v}_1^\top \mathbf{x}_t'$
Say \( w_1 \) and \( v_1 \) are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

\[
\frac{1}{n} \sum_{t=1}^{n} \left( y_t[1] - \frac{1}{n} \sum_{t=1}^{n} y_t[1] \right) \cdot \left( y_t'[1] - \frac{1}{n} \sum_{t=1}^{n} y_t'[1] \right)
\]

s.t. \( \frac{1}{n} \sum_{t=1}^{n} \left( y_t[1] - \frac{1}{n} \sum_{t=1}^{n} y_t[1] \right)^2 = \frac{1}{n} \sum_{t=1}^{n} \left( y_t'[1] - \frac{1}{n} \sum_{t=1}^{n} y_t'[1] \right) = 1 \)

where \( y_t[1] = w_1^\top x_t \) and \( y_t'[1] = v_1^\top x_t' \)
Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

maximize $\frac{1}{n} \sum_{t=1}^{n} \mathbf{w}_1^\top (\mathbf{x}_t - \mu) \cdot \mathbf{v}_1^\top (\mathbf{x}_t' - \mu')$

subject to $\frac{1}{n} \sum_{t=1}^{n} (\mathbf{w}_1^\top (\mathbf{x}_t - \mu))^2 = \frac{1}{n} \sum_{t=1}^{n} (\mathbf{v}_1^\top (\mathbf{x}_t' - \mu'))^2 = 1$

where $\mu = \frac{1}{n} \sum_{t=1}^{n} \mathbf{x}_t$ and $\mu' = \frac{1}{n} \sum_{t=1}^{n} \mathbf{x}_t'$
Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

maximize $\mathbf{w}_1^\top \Sigma_{1,2} \mathbf{v}_1$

subject to $\mathbf{w}_1^\top \Sigma_{1,1} \mathbf{w}_1 = \mathbf{v}_1^\top \Sigma_{2,2} \mathbf{v}_1 = 1$
Hence we want to solve for projection vectors $w_1$ and $v_1$ that maximize $w_1^\top \Sigma_{1,2} v_1$

subject to $w_1^\top \Sigma_{1,1} w_1 = v_1^\top \Sigma_{2,2} v_1 = 1$
\[ W = \text{eigs}(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}, K) \]

\[ V = \text{eigs}(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}, K) \]
Write \( \tilde{x}_t = x_t x_t' \) the \( d + d' \) dimensional concatenated vectors. Calculate covariance matrix of the joint data points \( \Sigma = \begin{pmatrix} \Sigma_1, 1 & \Sigma_1, 2 \\ \Sigma_2, 1 & \Sigma_2, 2 \end{pmatrix} \). The top \( K \) eigen vectors of this matrix give us projection matrix for view I. Calculate \( \Sigma_{-1}, 1, 1 \Sigma_{1, 1}, 2 \Sigma_{-1, 2, 1} \Sigma_{1, 2, 2} \). The top \( K \) eigen vectors of this matrix give us projection matrix for view II.
1. \( X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \)
Write $\tilde{x}_t = x_t x'_t$ the $d + d'$ dimensional concatenated vectors.

Calculate covariance matrix of the joint data points

$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \text{cov}(X)$

The top $K$ eigen vectors of this matrix give us projection matrix for view I.

Calculate $\Sigma_{12}$, $\Sigma_{22}$, $\Sigma_{21}$, $\Sigma_{11}$. The top $K$ eigen vectors of this matrix give us projection matrix for view II.
1. \( X = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \)

2. \( \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \text{cov}(X) \)

3. \( W = \text{eigs}(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}, K) \)
CCA Algorithm

1. \( X = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \)

2. \( \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \text{cov}(X) \)

3. \( W = \text{eigs}(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, K) \)

4. \( Y_1 = (X_1 - \mu_1) \times W \)