Applied Machine Learning

Canonical Correlation Analysis

Canonical Correlation Analysis

- Overview
- Projections of item pairs
- Maximizing correlations
- Significance of correlations

Canonical Correlation Analysis

- Associate pairs of correlated items from two different datasets
    - Projections onto vectors that maximize correlation
- Applications
    - Images with text descriptions
    - Segments of videos and corresponding captions
    - Segments of accelerometer signals and description of user activity

CCA: Projections

- Dataset $\{\mathbf{p}\}$
  - Data item i: $\mathbf{p}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{bmatrix}$
    - $\mathbf{x}_i \in \{\mathbf{x}\}$
      - d_x features, mean$(\{\mathbf{x}\}) = 0$
    - $\mathbf{y}_i \in \{\mathbf{y}\}$
      - d_y features, mean$(\{\mathbf{y}\}) = 0$

- Projection of $\{\mathbf{x}\}$ onto $\mathbf{u}$
  - $\{\mathbf{u}^\top \mathbf{x}\}$, projection i: $\{\mathbf{u}^\top \mathbf{x}_i\}$
- Projection of $\{\mathbf{y}\}$ onto $\mathbf{v}$
  - $\{\mathbf{v}^\top \mathbf{y}\}$, projection i: $\{\mathbf{v}^\top \mathbf{y}_i\}$
- Goal, find $\mathbf{u}, \mathbf{v}$ that maximize:
  - corr$(\{\mathbf{u}^\top \mathbf{x}, \mathbf{v}^\top \mathbf{y}\})$

Maximizing Correlation of Projections

- Covariance Matrix of $\{\mathbf{p}\}$
- $\Sigma = \begin{bmatrix} \Sigma_{x,x} & \Sigma_{x,y} \\ \Sigma_{y,x} & \Sigma_{y,y} \end{bmatrix}$

$$\text{corr}(\{(x, y)\}) = \frac{\text{cov}(\{x\}, \{y\})}{\text{std}(x)\text{std}(y)}$$

- $\text{corr}(\{\mathbf{u}^\top \mathbf{x}, \mathbf{v}^\top \mathbf{y}\}\}) = \dfrac{\mathbf{u}^\top \Sigma_{x,y} \mathbf{v}}{\sqrt{\mathbf{u}^\top \Sigma_{x,x} \mathbf{u}} \sqrt{\mathbf{v}^\top \Sigma_{y,y} \mathbf{v}}}$

- Lagrange multiplier:
- max$(\mathbf{u}^\top \Sigma_{x,y} \mathbf{v})$ subject to $\mathbf{u}^\top \Sigma_{x,x} \mathbf{u} = c_1$ and $\mathbf{v}^\top \Sigma_{y,y} \mathbf{v} = c_2$

- $\mathbf{u}^\top \Sigma_{x,y} \mathbf{v} - \lambda_1(\mathbf{u}^\top \Sigma_{x,x} \mathbf{u} - c_1) - \lambda_2(\mathbf{v}^\top \Sigma_{y,y} \mathbf{v} - c_2)$
- Solve:
- $\Sigma_{x,y} \mathbf{v} - \lambda_1 \Sigma_{x,x} \mathbf{u} = 0 \quad \Sigma_{x,y}^\top \mathbf{u} - \lambda_2 \Sigma_{y,y} \mathbf{v} = 0$
  - $\Sigma_{x,x}$ and $\Sigma_{y,y}$: invertible
- $\mathbf{u} = \dfrac{1}{\lambda_1} \Sigma_{x,x}^{-1} \Sigma_{x,y} \mathbf{v} \quad \mathbf{v} = \dfrac{1}{\lambda_2} \Sigma_{y,y}^{-1} \Sigma_{x,y}^\top \mathbf{u}$
- $\lambda_1 \lambda_2 \mathbf{u} = \Sigma_{x,x}^{-1} \Sigma_{x,y} \Sigma_{y,y}^{-1} \Sigma_{x,y}^\top \mathbf{u} \quad \lambda_1 \lambda_2 \mathbf{v} = \Sigma_{y,y}^{-1} \Sigma_{x,y}^\top \Sigma_{x,x}^{-1} \Sigma_{x,y} \mathbf{v}$
  - $\mathbf{u}$: eigenvector of $\Sigma_{x,x}^{-1} \Sigma_{x,y} \Sigma_{y,y}^{-1} \Sigma_{x,y}^\top$
  - $\mathbf{v}$: is eigenvector of $\Sigma_{y,y}^{-1} \Sigma_{x,y}^\top \Sigma_{x,x}^{-1} \Sigma_{x,y}$

Maximizing Correlation of Projections

$\mathbf{u}^\top\Sigma_{x,y}\mathbf{v} = \mathbf{u}^\top(\lambda_1\Sigma_{x,x}\mathbf{u}) \qquad \mathbf{u}^\top\Sigma_{x,y}\mathbf{v} = \mathbf{v}^\top(\lambda_2\Sigma_{y,y}\mathbf{v})$

- $\sqrt{\frac{\mathbf{u}^\top\Sigma_{x,y}\mathbf{v}}{\lambda_1\mathbf{u}^\top\Sigma_{x,x}\mathbf{u}}} = 1 \qquad \sqrt{\frac{\mathbf{u}^\top\Sigma_{x,y}\mathbf{v}}{\lambda_2\mathbf{v}^\top\Sigma_{y,y}\mathbf{v}}} = 1$
- $\frac{\mathbf{u}^\top\Sigma_{x,y}\mathbf{v}}{\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\mathbf{u}^\top\Sigma_{x,x}\mathbf{u}}\sqrt{\mathbf{v}^\top\Sigma_{y,y}\mathbf{v}}} = 1$
- $\frac{\mathbf{u}^\top\Sigma_{x,y}\mathbf{v}}{\sqrt{\mathbf{u}^\top\Sigma_{x,x}\mathbf{u}}\sqrt{\mathbf{v}^\top\Sigma_{y,y}\mathbf{v}}} = \sqrt{\lambda_1}\sqrt{\lambda_2} = \text{corr}(\{\mathbf{u}^\top\mathbf{x}, \mathbf{v}^\top\mathbf{y}\})$
- Eigenvectors $\mathbf{u}$ and $\mathbf{v}$: descending order of their eigenvalues
  - $\mathbf{u}_1^\top\mathbf{x}_i$ has the strongest correlation with $\mathbf{v}_1^\top\mathbf{y}_i$
  - $\mathbf{u}_2^\top\mathbf{x}_i$ has the second strongest correlation with $\mathbf{v}_2^\top\mathbf{y}_i$

- $\Sigma_{x,y}\mathbf{v} - \lambda_1\Sigma_{x,x}\mathbf{u} = 0 \qquad \Sigma_{x,y}^\top\mathbf{u} - \lambda_2\Sigma_{y,y}\mathbf{v} = 0$
- $\mathbf{u} = \frac{1}{\lambda_1}\Sigma_{x,x}^{-1}\Sigma_{x,y}\mathbf{v} \qquad \mathbf{v} = \frac{1}{\lambda_2}\Sigma_{y,y}^{-1}\Sigma_{x,y}^\top\mathbf{u}$
- $\lambda_1\lambda_2\mathbf{u} = \Sigma_{x,x}^{-1}\Sigma_{x,y}\Sigma_{y,y}^{-1}\Sigma_{x,y}^\top\mathbf{u} \qquad \lambda_1\lambda_2\mathbf{v} = \Sigma_{y,y}^{-1}\Sigma_{x,y}^\top\Sigma_{x,x}^{-1}\Sigma_{x,y}\mathbf{v}$
- number of directions: $\min(d_x, d_y)$
- Canonical correlations: $\text{corr}(\{\mathbf{u}^\top\mathbf{x}, \mathbf{v}^\top\mathbf{y}\})$
- Canonical variables: projections

Significance of Correlations

- For $(r = 1; r < = \min(d_x, d_y); r + +)$
  - Compute Wilk's lambda of r:
    $$\Lambda(r) = \prod_{i=1}^{r} (1 - \rho_i^2),$$
    - with ρ_i the i'th canonical correlation
- The largest the correlation, the smallest the value of $\Lambda(r)$
- As r increases, $\Lambda(r)$ reduces

- Apply to several permutations of data $\{\mathbf{y}\}$
- if $\Lambda(r)$ smaller than most of the ones for permuted datasets
  - more likely that correlations are meaningful
- Use numerical libraries

Canonical Correlation Analysis

- Overview
- Projections of item pairs
- Maximizing correlations
- Significance of correlations

Applied Machine Learning

Canonical Correlation Analysis