# Low Rank Approximations, PCA, and CCA

Abhinav Gopal, Shrey Vasavada, William McEachen, Grace Kull, Jai Bansal



# LOW RANK APPROXIMATIONS | EECS 189

- What is a low rank approximation?
  - Large amounts of data is expensive
  - Furthermore, some \*unimportant\* aspects of data reflect noise
  - How do we create an approximation of the data with just the important features?
- Option: Reduce features
  - How? Choose the most "important" features of the matrix
  - Decompose to svd and only use the top "k" values
  - $\circ \quad \boldsymbol{\Sigma}_{i \in (1...k)} \boldsymbol{u}_i \boldsymbol{\sigma}_i \boldsymbol{v}^T_{\ i}$



# LOW RANK APPROXIMATIONS | EECS 189

- Notice from the previous slide that our low rank approximation's column space was the k vectors v<sub>i</sub>
- These vectors represent the directions of most variance in our data
- Idea: Use these vectors in Xw = y prediction tasks
  - Project X onto these important directions, and use this projection to predict y.



### **PCA** | **EECS** 189

- This brings us to PCA!
- Procedure
  - Compute SVD for matrix X
  - Get the top k eigenvectors V<sup>T</sup><sub>i</sub> and choose these eigenvectors
  - Get the principal components by finding Z<sub>k</sub> = XV<sub>k</sub>
  - Get the approximation for X by doing ZV<sup>T</sup><sub>k</sub>
  - Use this approximation and do least squares



### **PCA** | **EECS** 189

- Great, but flawed, way to approximate a matrix and eliminate noise.
  - Why? Relies on the assumption that the largest variance directions are the most important ones.
- Very easy to modify data in order to render PCA useless
- What is the workaround?
  - We try to find the actual correlation between X and y.



- This brings us to CCA
  - Canonical Correlation Analysis
- CCA uses the pearson correlation coefficient in order to determine the relationship between X and y.

$$\hat{\rho}(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$= \frac{\tilde{x}^T \tilde{y}}{\sqrt{\tilde{x}^T \tilde{x} \cdot \tilde{y}^T \tilde{y}}} \quad \text{where } \tilde{x} = x - \bar{x}, \tilde{y} = y - \bar{y}$$



Goal: To Solve

$$\max_{\mathbf{u},\mathbf{v}} \rho(\mathbf{X}_{rv}^{\top}\mathbf{u}, \mathbf{Y}_{rv}^{\top}\mathbf{v}) = \max_{\mathbf{u},\mathbf{v}} \frac{\text{Cov}(\mathbf{X}_{rv}^{\top}\mathbf{u}, \mathbf{Y}_{rv}^{\top}\mathbf{v})}{\sqrt{\text{Var}(\mathbf{X}_{rv}^{\top}\mathbf{u}) \text{Var}(\mathbf{Y}_{rv}^{\top}\mathbf{v})}}$$

- X<sub>rv</sub> and Y<sub>rv</sub> are both real vectors that correspond to the size of X and Y.
- With some algebra, we find that the problem becomes

$$= \max_{\mathbf{u}, \mathbf{v}} \frac{\mathbf{u}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{Y} \mathbf{v}}{\sqrt{\mathbf{u}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{X} \mathbf{u} \cdot \mathbf{v}^{\mathsf{T}} \mathbf{Y}^{\mathsf{T}} \mathbf{Y} \mathbf{v}}}$$

$$\hat{\rho}(\mathbf{X} \mathbf{u}, \mathbf{Y} \mathbf{v})$$



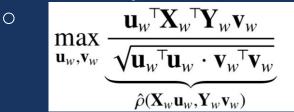
Whiten the X and Y matrices (make their covariance 0) by

$$\circ$$
 Use W<sub>x</sub> = U<sub>x</sub>S<sup>1/2</sup><sub>x</sub>U<sub>x</sub><sup>T</sup>

- $O Use W_y = U_y S^{1/2}_y U_y$
- $\circ$  SVD of X<sup>T</sup>X =  $U_x S_x V_x^T$
- $\circ$  SVD of Y<sup>T</sup>Y =  $U_y S_y V_y^T$
- $\circ$  Use  $X_w = XW_x, Y_w = YW_Y$



 With the change of variables u<sub>w</sub> = W<sub>x</sub><sup>-1</sup>u and v<sub>w</sub> = W<sub>y</sub><sup>-1</sup>v, the max expression becomes:



- Now, let's decorrelate  $X_w$  and  $Y_w$ , making  $(X_wD_x)^T(Y_wD_v)$
- We decorrelate them with the choice of
  - $\circ$  U for  $D_x$  and V for  $D_y$ , where  $USV^T = X_w^T Y_w$



- Changing variables with
  - $\circ$   $u_d = D_x^T u_w$  and  $v_d = D_y^T v_w$ , the maximization becomes the following:

$$\max_{\substack{\|\mathbf{u}_d\|=1\\\|\mathbf{v}_d\|=1}}\mathbf{u}_d^\mathsf{T}\mathbf{S}\mathbf{v}_d$$



Finally, we get the associated eigenvectors of interest!  $\circ$  U = W<sub>x</sub>D<sub>x</sub>U<sub>d</sub>, V = W<sub>y</sub>D<sub>y</sub>V<sub>d</sub>, with

$$\mathbf{U}_d = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}_{p-k,k} \end{bmatrix} \in \mathbb{R}^{p \times k}$$

$$\mathbf{U}_d = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}_{p-k,k} \end{bmatrix} \in \mathbb{R}^{p \times k} \qquad \qquad \mathbf{V}_d = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}_{q-k,k} \end{bmatrix} \in \mathbb{R}^{q \times k}$$

