Admin

- Lab 5 grades posted
- No lab on Nov. 5
- Prof Mathieson will cover Nov. 6 lecture

Dimensionality Reduction - used often for data visualization

- Principal Component Analysis (PCA) - algorithm that transforms high-dimensional data to usually 2 dimensional, or lower dimensional data
- Helps us see relationships that might not be obvious in high-dimensional form

Global Population structure example

- Similar ethnic groups can be seen in clusters based on principal component 1 and principal component 2 in 2 dimensions
- Populations splits separate populations
- Admixture merges populations
- Principal component 1 captures the biggest split in the data

PCA application: Eigenfaces

- Faces are inherently high-dimensional
  - You have all the pixels and their RGB values
- Can use PCA to lower the dimension to capture different features in the human face
- Can be combined with to rebuild the original image
- Compare how faces are represented by the basis set
  - Weighted representation of the number of eigenfaces
- All the features are distinct and represent different parts of the face

PCA for data visualization:

- Step 1: original matrix is n x p
  - Where p is much larger than n
  - Goal: create n x 2 matrix for visualization
- Step 2:
  - Subtract off column-wise mean

$$\begin{bmatrix} 2 & 1 \\ 3 & 3 \end{bmatrix}$$

$\bar{x}_1 = 2.5 \quad \bar{x}_2 = 2$

$$\begin{bmatrix} 2-2.5 & 1-2 \\ 3-2.5 & 3-2 \end{bmatrix}$$

$$\begin{bmatrix} -.5 & -1 \\ .5 & 1 \end{bmatrix}$$

  -

- Step 3:
    - Compute covariance matrix A
    - Same idea for multiple features

Step 3

f, g, h

$$A = \begin{bmatrix} cov(f,f) & cov(f,h) & cov(f,g) \\ \vdots & & \end{bmatrix}$$

p x p

        -
    - Runtime to calculate a pair of features is O(n)
        - To compute covariance matrix: O(np^2)
- Step 4:
    - Compute eigenvalues and eigenvalues of A
        - v is being scaled, but direction is not changing

Step 4

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

    -
    - Principle components that we end up with are basically the eigenvectors
- Step 5:
    - Sort eigenvectors by eigenvalues (high->low)
    - Compute transformed data:
        - $T_{nxr} = X_{nxp} W_{pxr}$