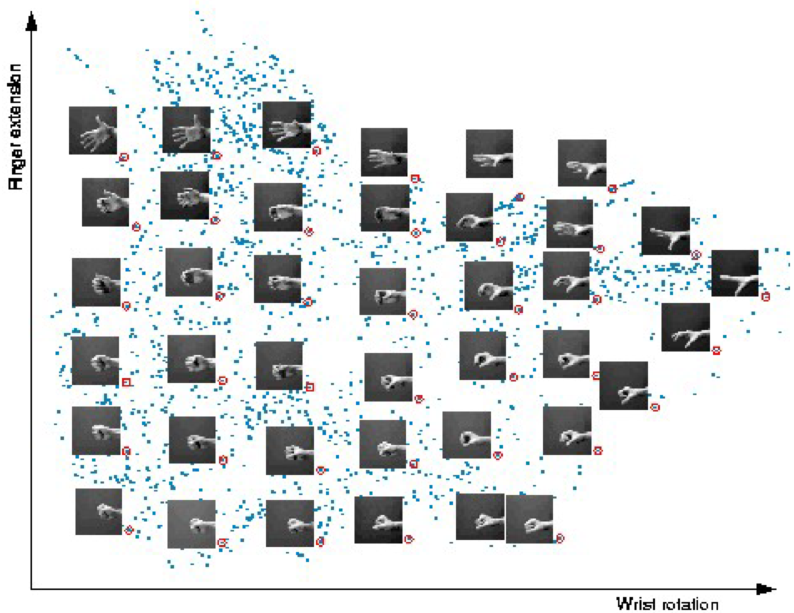
# Dimensionality Reduction:
# Principal Component Analysis (PCA)
# & Singular Value Decomposition (SVD)

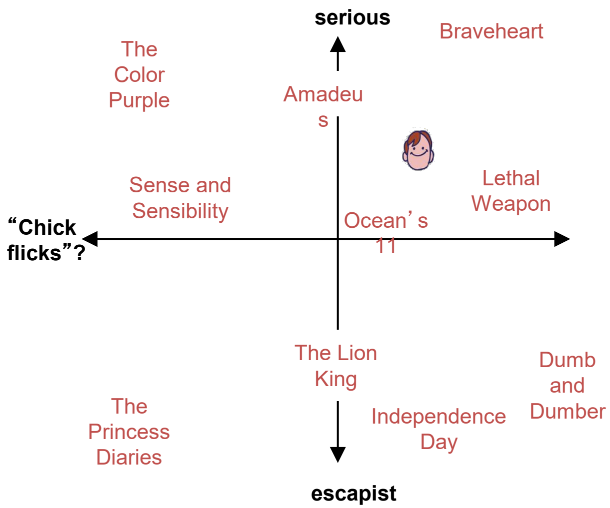
# Motivation

- High-dimensional data
  - Images of faces
  - Text from articles
  - All S&P 500 stocks
- Can we describe them in a "simpler" way?
  - Embedding: place data in $R^d$, such that "similar" data are close

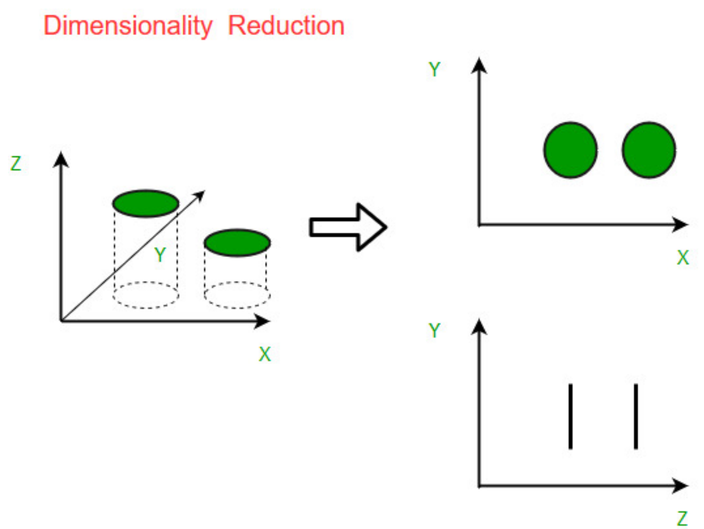
Ex: embedding images in 2D

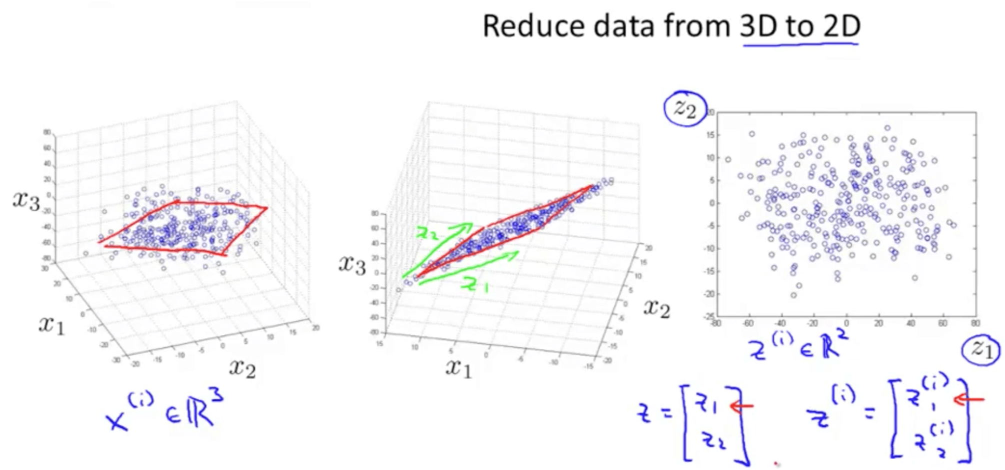



Ex: embedding movies in 2D



# Dimensionality Reduction

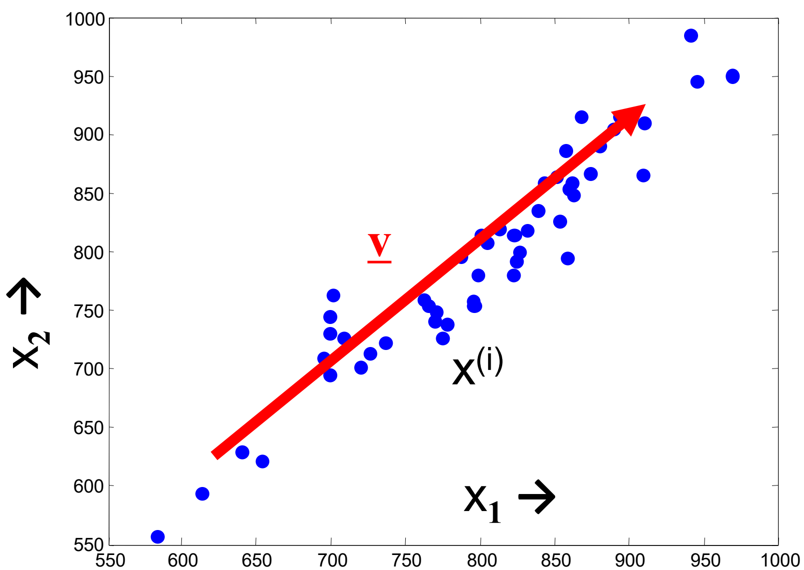


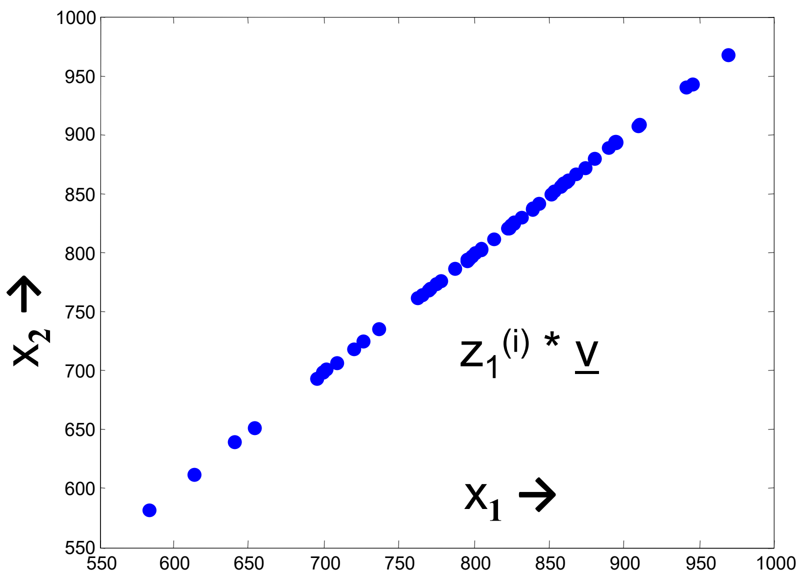
# Dimensionality reduction

- Ex: data with two real values $[x_1,x_2]$
- We'd like to describe each point using only one value $[z_1]$
- We'll communicate a "model" to convert: $[x_1,x_2] \sim f(z_1)$

- Ex: linear function f(z): $[x_1,x_2] = z_1 * \underline{v} = z_1 * [v_1,v_2]$
- $\underline{v}$ is the same for all data points (communicate once)
- z tells us the closest point on v to the original point $[x_1,x_2]$



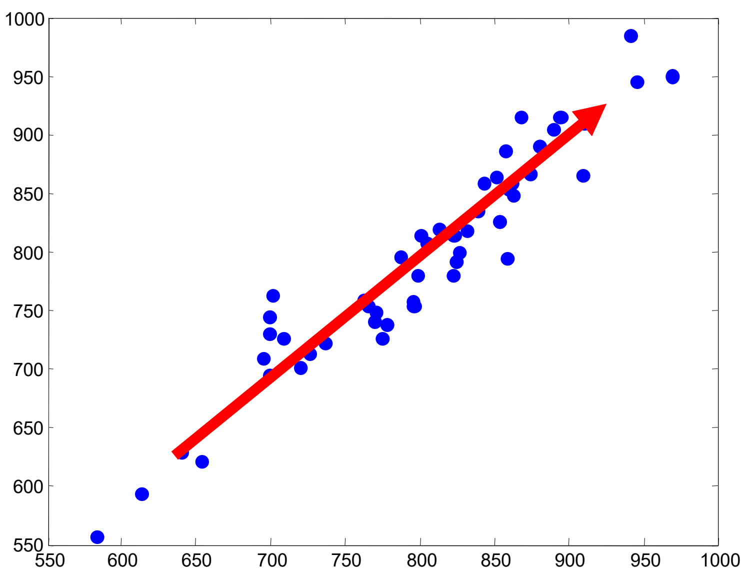

Project x to v: $z_1 = x \cdot v$



# Principal Components Analysis

- How should we find v?
  - Assume X is zero mean, or $\tilde{X} = X - \mu$
  - Find "v" as the direction of maximum "spread" (variance)
  - Solution is the eigenvector with largest eigenvalue

  - Equivalent: v also leaves the smallest residual variance! ("error")




Project X to v: $z = \tilde{X} \cdot v$

Variance of projected points:

$$\sum_i (z^{(i)})^2 = z^T z = v^T \tilde{X}^T \tilde{X} v$$

Best "direction" v:

$$\max_v \; v^T \tilde{X}^T \tilde{X} v \quad s.t.\ ||v|| = 1$$

- largest eigenvector of $X^T X$

# Eigenvector Decomposition

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Oval shows constant $\Delta^2$ value…

$$\Sigma = U \Lambda U^T$$

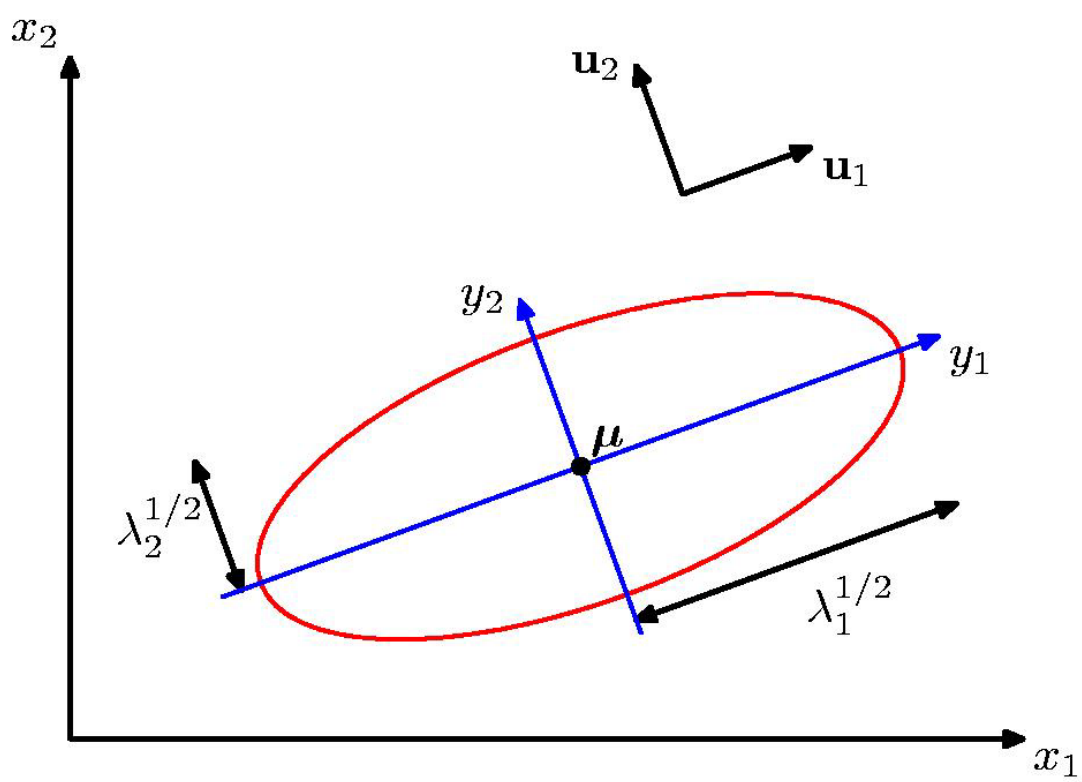
Write Σ in terms of eigenvectors…

$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

Then…

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

$$y_i = \mathbf{u}_i^{\mathrm{T}} (\mathbf{x} - \boldsymbol{\mu})$$



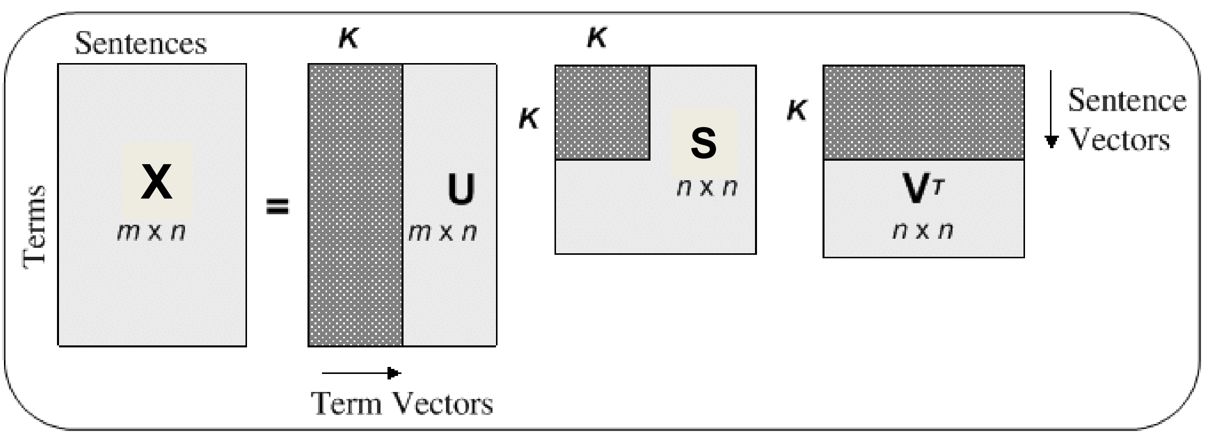
# PCA representation

- Subtract data mean from each point
- (Typically) scale each dimension by its variance
  - Helps pay less attention to magnitude of the variable
- Compute covariance matrix, $\Sigma = 1/m \sum (x^i - {}^1)' \; (x^i - {}^1)$
- Compute the k largest eigenvectors of Σ

  $\Sigma = V\ D\ V\hat{}T$
- Compute the projections onto k eigenvectors for all data examples.

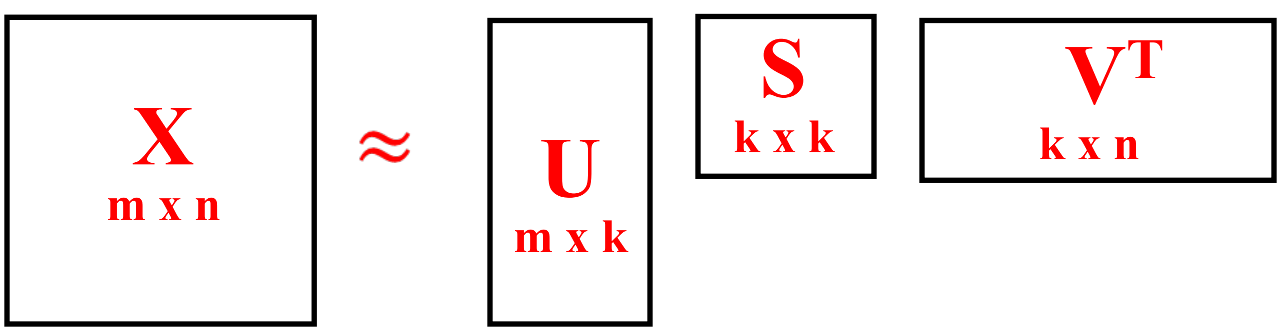
# Singular Value Decomposition

- Alternative method to calculate (still subtract mean 1st)
- Decompose $X = U\,S\,V^T$
  - Orthogonal: $X^T X = V\,S\,S\,V^T = V\,D\,V^T$
- U*S matrix provides coefficients
  - Example $x^{(i)} = U_{i,1}\,S_{11}\,v_1 + U_{i,2}\,S_{22}\,v_2 + \ldots$
- Gives the least-squares approximation to X of this form



# Singular Value Decomposition

- Alternative method to calculate (still subtract mean 1st)
- Decompose $X = U\,S\,V^T$
  - Orthogonal: $X^T X = V\,S\,S\,V^T = V\,D\,V^T$
- U*S matrix provides coefficients
  - Example $x^{(i)} = U_{i,1}\,S_{11}\,v_1 + U_{i,2}\,S_{22}\,v_2 + \ldots$
- Gives the least-squares approximation to X of this form



# SVD for PCA

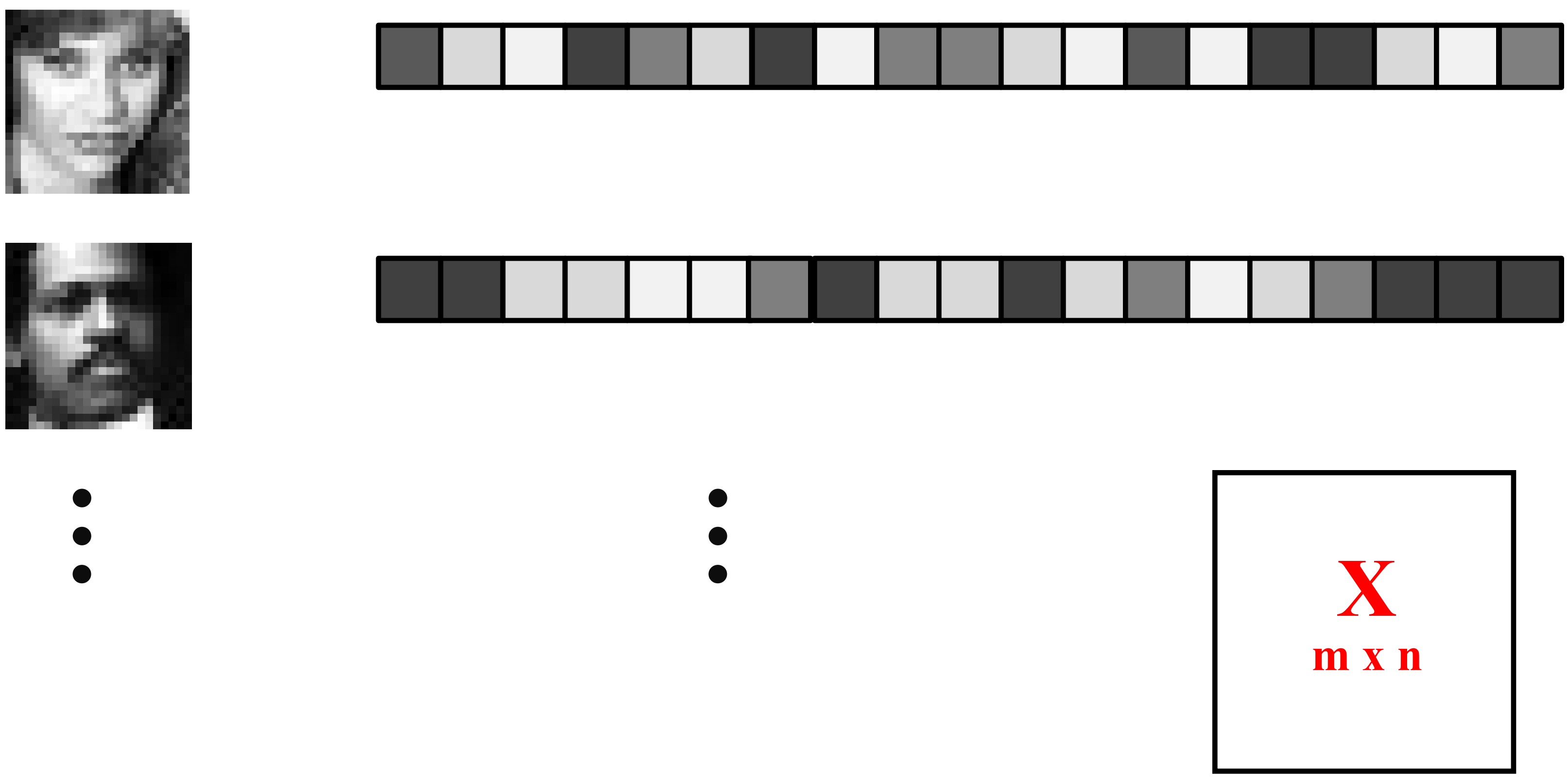
- Subtract data mean from each point
- (Typically) scale each dimension by its variance
  - Helps pay less attention to magnitude of the variable
- Compute the SVD of the data matrix

  $X = U S V^T$
- Extract first k columns in US

# APPLICATIONS OF PCA/SVD

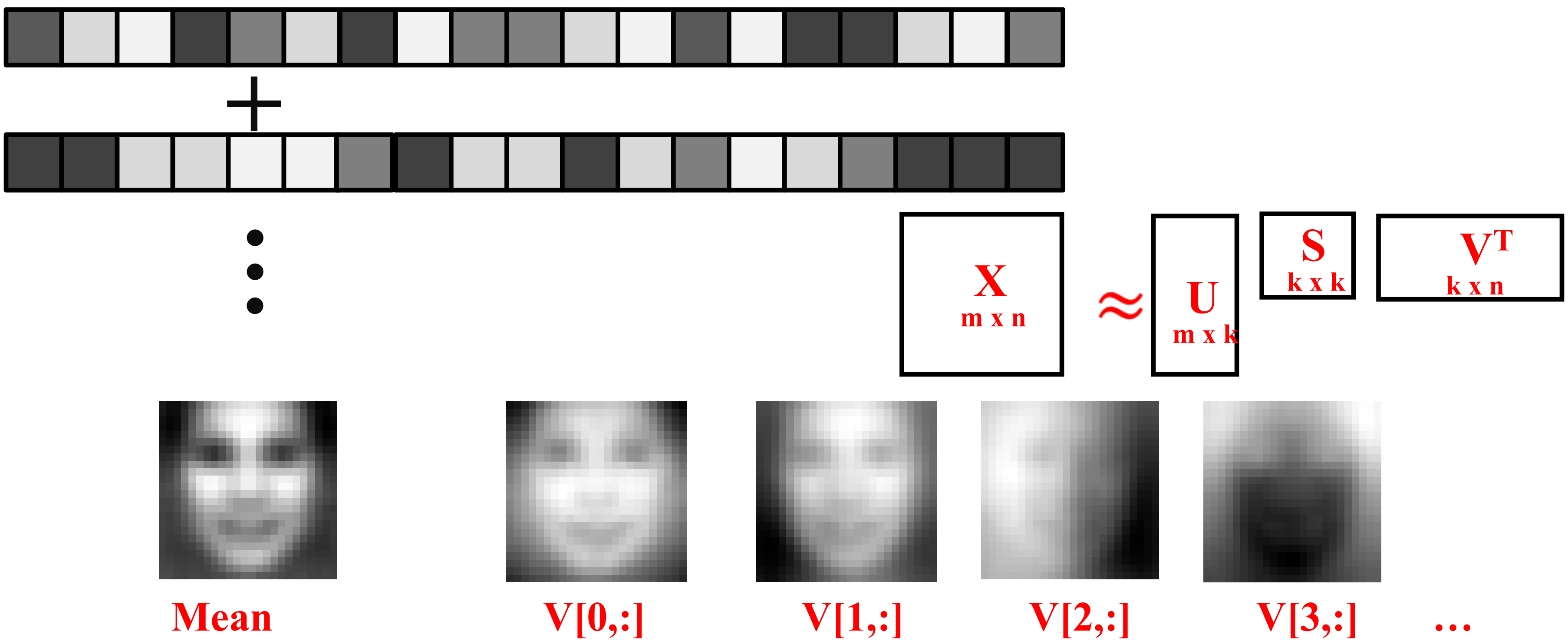
# “Eigen-faces”

- “Eigen-X” = represent X using PCA
- Ex: Viola Jones data set
  - 24x24 images of faces = 576 dimensional measurements



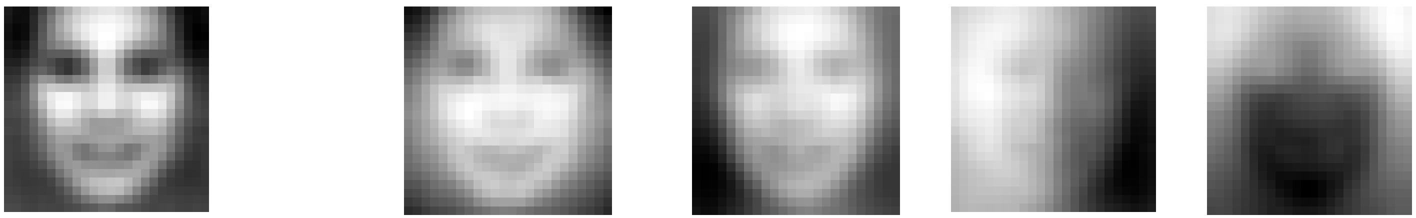
# “Eigen-faces”

- “Eigen-X” = represent X using PCA
- Ex: Viola Jones data set
  - 24x24 images of faces = 576 dimensional measurements
  - Take first K PCA components



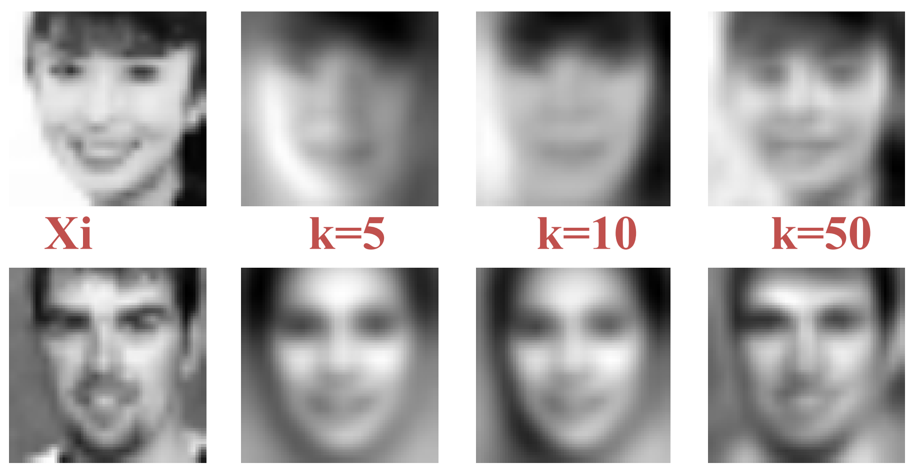
# "Eigen-faces"

- "Eigen-X" = represent X using PCA
- Ex: Viola Jones data set
  - 24x24 images of faces = 576 dimensional measurements
  - Take first K PCA components



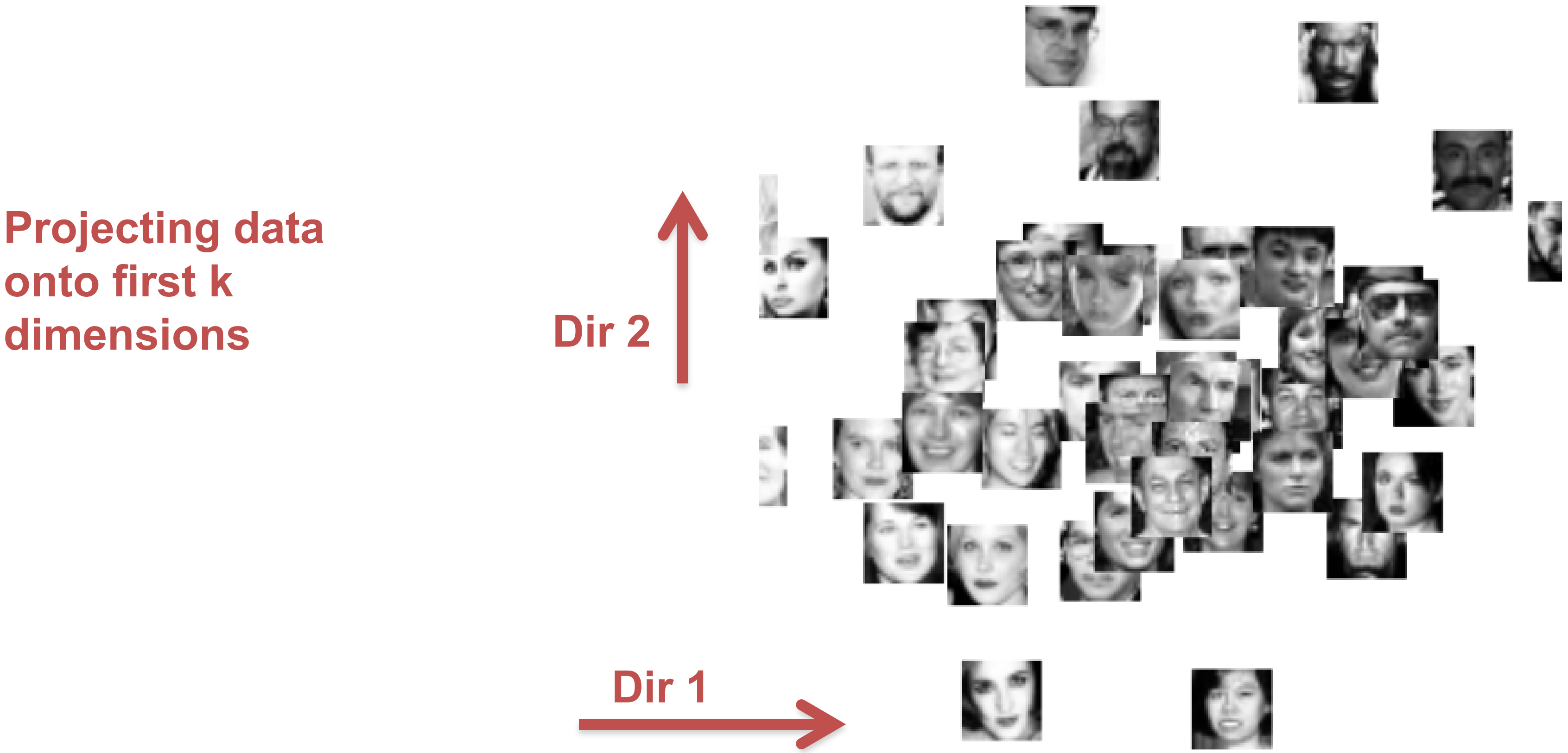
Mean Dir 1 Dir 2 Dir 3 Dir 4 …

Projecting data onto first k dimensions



# "Eigen-faces"

- "Eigen-X" = represent X using PCA
- Ex: Viola Jones data set
  - 24x24 images of faces = 576 dimensional measurements
  - Take first K PCA components



# Summary

- Dimensionality reduction
  - Representation: basis vectors & coefficients

- Linear decomposition
  - PCA / eigendecomposition
  - Singular value decomposition

- Examples and data sets
  - Face images
  - Text representation