Dimensionality reduction

- Input data may have thousands or millions of dimensions!
  - e.g., text data has
  - $\text{10,000} \rightarrow \text{10,000,000}$

- **Dimensionality reduction**: represent data with fewer dimensions
  - easier learning – fewer parameters
  - visualization – hard to visualize more than 3D or 4D
  - discover “intrinsic dimensionality” of data
  - high dimensional data that is truly lower dimensional
Lower dimensional projections

- Rather than picking a subset of the features, we can create new features that are combinations of existing features.

\[ z = 2.5x_1 + 2.9x_2 - 3.2x_3 \]

- Let’s see this in the unsupervised setting.
  - Only \( X \), but no \( Y \).

Linear projection and reconstruction

- Project into 1-dimension.
- Reconstruction: only know \( z_1 \), how far from original point.
- Reconstruction error: overlap between \( z \), and \( X \).
Principal component analysis – basic idea

- Project d-dimensional data into k-dimensional space while preserving information:
  - e.g., project space of 10,000 words into 3-dimensions
  - e.g., project 3-d into 2-d

- Choose projection with minimum reconstruction error

Linear projections, a review

- Project a point into a (lower dimensional) space:
  - point: \( \mathbf{x} = (x_1, \ldots, x_d) \)
  - select a basis – set of basis vectors – \((u_1, \ldots, u_k)\)
    - we consider orthonormal basis:
      - \( u_i \cdot u_i = 1 \), and \( u_i \cdot u_j = 0 \) for \( i \neq j \)
  - select a center – \( \bar{x} \), defines offset of space
  - best coordinates in lower dimensional space defined by dot-products: \((z_1, \ldots, z_k)\), \( z_i = (\mathbf{x} - \bar{x}) \cdot u_i \)
    - minimum squared error

\[ z_i = (\mathbf{x} - \bar{x}) \cdot u_i \]

\[ z_i = \arg \min \frac{1}{2} (\mathbf{x} - \bar{x}) \cdot (\mathbf{x} - \bar{x}) - z_i \cdot u_i \]
PCA finds projection that minimizes reconstruction error

- Given \( k \) data points: \( \mathbf{x}^i = (x_1^i, \ldots, x_d^i) \), \( i=1\ldots N \)
- Will represent each point as a projection:
  \[
  \hat{x}^i = \bar{x} + \sum_{j=1}^{k} z_j^i \mathbf{u}_j
  \]
where: \( \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x^i \), and \( z_j^i = (x^i - \bar{x}) \cdot \mathbf{u}_j \)

PCA:
- Given \( k<d \), find \((\mathbf{u}_1, \ldots, \mathbf{u}_k)\)
mimizing reconstruction error:
  \[
  \text{error}_k = \sum_{i=1}^{N} (x^i - \hat{x}^i)^2
  \]

Understanding the reconstruction error

- Note that \( \mathbf{x}^i \) can be represented exactly by \( d \)-dimensional projection:
  \[
  \hat{x}^i = \bar{x} + \sum_{j=1}^{d} z_j^i \mathbf{u}_j
  \]

- Rewriting error:

\[
\text{error}_k = \sum_{i=1}^{N} (x^i - \hat{x}^i)^2 = \sum_{i=1}^{N} \left[ \frac{1}{N} \sum_{j=1}^{d} \left( z_j^i \mathbf{u}_j \right)^2 \right] = \sum_{j=1}^{d} \left( \frac{1}{N} \sum_{i=1}^{N} z_j^i \right)^2
\]
Reconstruction error and covariance matrix

\[ \text{error}_k = \sum_{i=1}^{N} \sum_{j=k+1}^{d} [u_j \cdot (x_i - \bar{x})]^2 \]

\[ = \sum_{i=1}^{N} \sum_{j=k+1}^{d} u_j^T (x_i - \bar{x})^T (x_i - \bar{x}) u_j \]

\[ = \sum_{j=k+1}^{d} \sum_{i=1}^{N} u_j^T \left[ \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \right] u_j \]

\[ = N \sum_{j=k+1}^{d} u_j^T \Sigma u_j \]

Minimizing reconstruction error equivalent to picking (ordered) orthonormal basis \((u_1, \ldots, u_d)\) minimizing:

\[ \Sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i^T - \bar{x}) (x_i - \bar{x})^T \]

\[ \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{pmatrix} \]

\[ \Sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \]

Minimizing reconstruction error equivalent to picking (ordered) orthonormal basis \((u_1, \ldots, u_d)\) minimizing:

\[ \sum_{k=1}^{d} \sigma_k \]

Minimizing reconstruction error equivalent to picking \((u_{k+1}, \ldots, u_d)\) to be eigen vectors with smallest eigen values

\[ \min \text{error} = \text{truncating out d-k eigen vectors with smallest eigen values} \]

\[ \Rightarrow \text{keep top k eigen vectors of } \Sigma \]
Basic PCA algorithm

- Start from an m by n data matrix $X$
- Recenter: subtract mean from each row of $X$
  - $x_c \leftarrow x - \bar{x}$
  - $X_c = \{x - \bar{x}\}$
- Compute covariance matrix:
  - $\Sigma \leftarrow \frac{1}{N} X_c^T X_c$
- Find eigen vectors and values of $\Sigma$
- Principal components: k eigen vectors with highest eigen values

PCA example

$$\hat{x}^i = \bar{x} + \sum_{j=1}^{k} z_j^i u_j$$
PCA example – reconstruction

\[ \hat{x}^i = \bar{x} + \sum_{j=1}^{k} z_j^i u_j \]

only used first principal component

Eigenfaces [Turk, Pentland '91]

- Input images:
- Principal components:
Eigenfaces reconstruction

- Each image corresponds to adding 8 principal components:

  ![Image of eigenfaces reconstruction](image)

  "highly recognizable"