- Principal Component Analysis

- high dimensional data
  \[ x \in \mathbb{R}^d \quad (x_1, x_2, \ldots, x_d) \quad d \text{ large} \]

- image: \( 256 \times 256 \rightarrow 256 \times 256 \times 3 \) dimensional

- text: represent each word in vocabulary as a dimension
  values in vector = # times a word appear in the document.
  dimension of the vector = # words in vocabulary

- vectors in high dimensions often have correlations/redundancy
  - Principal Component Analysis: Technique to map high dimensional data to a small # of interesting directions.

- input: \( n \) data points \( x_1, x_2, \ldots, x_n \quad x_i \in \mathbb{R}^d \)

- output: \( u_1, u_2, \ldots, u_r \in \mathbb{R}^d \) \( r \) interesting directions

  - How to tell whether a direction is interesting

  - Idea: the direction where data has largest variance.

  - First direction: \( u \)
    \[ \max \frac{1}{n} \sum_{i=1}^{n} <x_i, u>^2 \quad \|u\|_2 = 1 \]

    \[ \sqrt{\frac{1}{d} \sum_{i=1}^{d} (u_i)^2} \quad \text{“variance” of data in direction} \]

    \[ \sim \text{“interesting” if variance is large} \]

  - Second direction: \( u_2 \)
    \[ \max \frac{1}{n} \sum_{i=1}^{n} <x_i, u>^2 \quad \|u\|_2 = 1 \]
- $k$-th direction $u_k$
  \[ \max_{||u||=1} \langle x_i, u \rangle^2 \]
  \[ \forall j \leq k \]
  \[ u_k \perp u_j \]

- Compute PCA directions

- **Power method.**
  \[ A = \sum_{i=1}^{n} x_i x_i^T = \begin{bmatrix} d & \vdots \\ \vdots & d \end{bmatrix} \]
  \[ d \frac{1}{d} \]

- **Recall:** eigenvalue and eigenvector
  \[ A v = \lambda v \]
  $v$ is a nonzero vector, then $\lambda$ is eigenvalue, $v$ eigenvector

- **Eigenvalue decomposition.** For any symmetric matrix $A$, can write
  \[ A = \sum_{i=1}^{d} \lambda_i u_i u_i^T \]
  \( \{\lambda_i\} \) are eigenvalues, \( \{u_i\} \) are eigenvectors.
  \( \{u_i\} \) are orthogonal to each other.

- **Claim:** In PCA, $u_1 = v_1$, $u_2 = v_2$, ..., $u_r = v_r$.

- **Power method**
  
  Initialize $u_0$ as a random vector.
  
  For $i = 1$ to $t$
  
  $u^i = A u^{i-1}$
  
  return $u_t / ||Au^{t-1}||$

  Observe $u^1 \sim A u^0$, $u^2 \sim A(Au^0) = A^2 u^0$

  $u^t \sim A^t u^0$

  **Claim:** With high probability, when $t \geq \left( \frac{\log \frac{d}{\epsilon}}{\lambda_2 - \lambda_1} \right)$, then $||u^t - v|| \leq \epsilon$.

  (to be proved in next lecture)