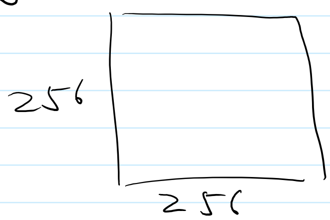
COMPSCI 330 Lecture 22 Basic Machine Learning

Wednesday, November 16, 2016 2:28 PM

- Principal Component Analysis
  - high dimensional data

    $x \in \mathbb{R}^d \quad (x_1, x_2, \ldots, x_d)$ d large

    - image: 256×256 ⟶ 256×256×3 dimensional



    - text: represent each word in vocabulary as a dimension

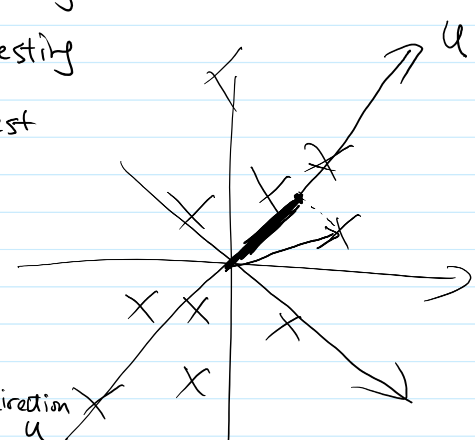
      values in vector = # times a word appear in the document.

      dimension of the vector = # words in vocabulary

    - vectors in high dimensions often have correlations/redundancy

- Principal Component Analysis: Techique to map high dimensional data to a small # of interesting directions.

  - input: n datapoints $x_1, x_2, \ldots, x_n \quad x_i \in \mathbb{R}^d$
  - 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_1

$$\max_{\|u\|=1} \sum_{i=1}^{n} \langle x_i, u \rangle^2$$

$$\sqrt{\sum_{i=1}^{d} (u(i))^2}$$

"variance" of data in direction u

~ "interesting" if variance is large

  - Second direction: u_2

$$\max_{\|u\|=1} \sum_{i=1}^{n} \langle x_i, u \rangle^2$$

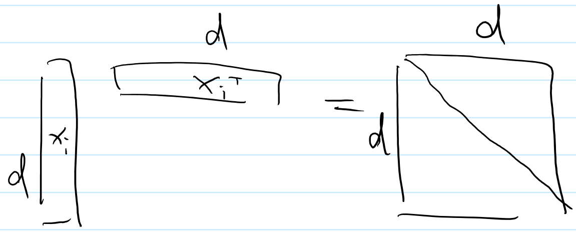
$u \perp u_1$

- k-th direction u_k

$$\max_{\substack{\|u\|=1 \\ \forall j<k \\ u_k \perp u_j}} \sum_{i=1}^n \langle x_i, u\rangle^2$$

- Compute PCA directions
  - Power method.

$$A = \sum_{i=1}^n x_i x_i^T = \sum_{i=1}^n$$



  - recall: eigenvalue and eigenvectors

$$A v = \lambda v$$

v is a nonzero vector, then λ is eigenvalue, v eigenvector

eigenvalue decomposition: for any symmetric matrix A, can write

$$A = \sum_{i=1}^d \lambda_i v_i v_i^T \qquad (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d)$$

$\{\lambda_i\}$'s are eigenvalues, $\{v_i\}$'s are eigenvectors,

$\{v_i\}$'s are orthogonal to each other.

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

  - power method

initialize u^0 as a random vector

for $i = 1$ to t

$$u^i = \frac{A u^{i-1}}{\|A u^{i-1}\|}$$

return u^t

observe $u^1 \sim A u^0 \quad u^2 \sim A(Au^0) = A^2 u^0$

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

  - Claim: With high probability, when $t \geq \frac{\left(\log \frac{d}{\varepsilon}\right)\lambda_1}{\lambda_1 - \lambda_2}$, then $\|u^t - v_1\| \leq \varepsilon$.

(to be proved in next lecture)