Tuesday, March 24

Welcome back!!

Plan

- Singular Value Decomposition
- Low-rank approxmation

Eigendecomposition

Square, symmetric $X \in \mathbb{R}^{d \times d}$

eigenvalue

$$Xv = \lambda v$$

eigenvector

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

$$v_1, v_2, \dots, v_d \in \mathbb{R}^d$$

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

$$X = \sum_{i=1}^{d} \lambda_i v_i v_i^T$$

$$XX =$$

$$X^T X =$$

Singular Vector Decomposition

Any matrix $X \in \mathbb{R}^{n \times d}$

WLOG, $n \geq d$

Singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d$

Left singular vectors $u_1, u_2, \dots, u_d \in \mathbb{R}^n$

Right singular vectors $v_1, v_2, \dots, v_d \in \mathbb{R}^d$

$$\langle u_i, u_j \rangle = \langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

$$X = \sum_{i=1}^{d} \sigma_i u_i v_i^T$$

$$X X =$$

$$X^T X =$$

$$Xa = \sum_{i=1}^{d} u_i \sigma_i v_i^T a$$

1. Project vector onto $v_1, \ldots, v_d$
2. Scale coordinates
3. Linear combination of $u_1, \ldots, u_d$

SVD useful for...

- Rank $\quad r = \{ i : \sigma_i > 0 \}$
- Pseudo inverse $\quad X^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$
- Condition number $\quad \sigma_1 / \sigma_d$
- Matrix norms i.e., $\|X\|_2 = \sigma_1$, $\|X\|_F^2 = \sum_{i=1}^{d} \sigma_i^2$
- Principal Component Analysis

Low Rank Approximation

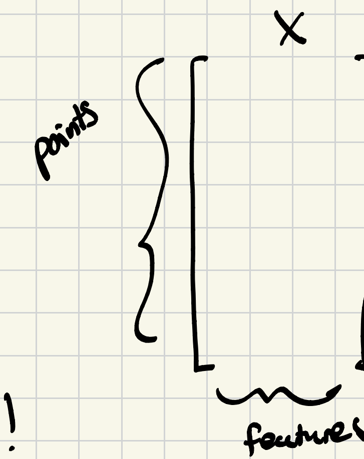
Rank characterizes structure:

↳ rank k if k unique points

↳ $\approx$ rank k if k clusters of points

↳ rank k if k indep. features

$\Rightarrow$ reduce dimension using structure!



Goal: Given X, best rank-k approximation

$$X_k = \underset{XWW^T \text{ where } W \in \mathbb{R}^{d \times k} \text{ with orthonormal columns ie } W^TW = I}{\operatorname{argmin}} \| X - XWW^T \|_F^2$$

Eckart-Young-Mirsky: $X_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$

$$(*) = \|X - XWW^T\|_F^2 = \|X(I - WW^T)\|_F^2$$

$$= tr\left((I-WW^T) X^T X (I-WW^T) \right)$$

$(I - WW^T) = (I-WW^T)^T = (I-WW^T)^2$

$$= tr\left(X^T X (I-WW^T)^2 \right)$$

cyclic property of the trace

$$= tr(X^T X) - tr(X^T X W W^T)$$

linearity of trace

$$\underset{W}{\arg\min}\ (*) = \underset{W}{\arg\max}\ tr(X^T X W W^T)$$

$$tr(X^T X W W^T) = tr(W^T X^T X W)$$

$$= tr\left(W^T \sum_{i=1}^{d} v_i \sigma_i^2 v_i^T W\right)$$

$$X^T X = \sum_{i=1}^{d} \sigma_i v_i u_i^T \sum_{j=1}^{d} \sigma_j u_j v_j^T$$

$$= \sum_{i=1}^{d} \sigma_i^2\ tr(W^T v_i v_i^T W)$$

$$= \sum_{i=1}^{d} \sigma_i^2 \|v_i^T W\|_2^2$$

$z_i = \|v_i^T W\|_2^2$ $\quad 0 \le z_i \le 1$ because columns of W are orthonormal

$$\sum_{i=1}^{d} z_i = \sum_{i=1}^{d} tr(W^T v_i v_i^T W) = tr(W^T \sum_{i=1}^{d} v_i v_i^T W)$$

$$= tr(W^T I_{d \times d} W) = tr(I_{k \times k}) = k$$

Choose z_i with $0 \le z_i \le 1$ and $\sum_{i=1}^{d} z_i = k$ to maximize $\sum_{i=1}^{d} \sigma_i^2 z_i$

$\therefore$ Choose $z_i = \begin{cases} 1 & \text{if } i \le k \\ 0 & \text{else} \end{cases}$. $W = \sum_{i=1}^{k} v_i e_i^T$

$$XWW^T = \sum_{i=1}^{d} \sigma_i u_i v_i^T \sum_{j=1}^{k} v_j e_j^T \sum_{l=1}^{k} e_l v_l^T$$

$$= \sum_{i=1}^{d} \sigma_i u_i v_i^T \sum_{j=1}^{k} v_j v_j^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

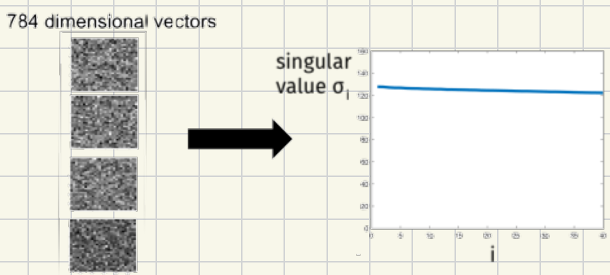
Q: How good is the approximation?

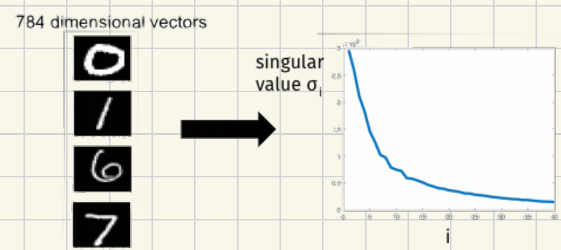
$$X - X_k = \sum_{i=1}^{d} \sigma_i u_i v_i^T - \sum_{i=1}^{k} \sigma_i u_i v_i^T = \sum_{i=k+1}^{d} \sigma_i u_i v_i^T$$

$$\|X - X_k\|_F^2 = tr\left((X-X_k)^T (X-X_k)\right) = tr\left(\sum_{j=k+1}^{d} \sigma_j v_j u_j^T \sum_{i=k+1}^{d} \sigma_i u_i v_i^T\right)$$

$$= tr\left(\sum_{i=k+1}^{d} \sigma_i^2 v_i v_i^T\right) = \sum_{i=k+1}^{d} tr\left(\sigma_i^2 v_i v_i^T\right)$$

$$= \sum_{i=k+1}^{d} \sigma_i^2 \, tr\left(v_i^T v_i\right) = \sum_{i=k+1}^{d} \sigma_i^2$$





Efficiently find SVD. Claim: When $n \geq d$, find $O(nd^2)$ time.

Hint: Getting eigendecomposition of $M \in \mathbb{R}^{m \times m}$ takes $O(m^3)$ time