Tuesday, March 24

Welcome back!!

Plan

- Singular Value Decomposition
- Low-rank approximation

Eigendecomposition

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

$Xv = \lambda v$ (λ: eigenvalue, v: eigenvector)

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$

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

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

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

$XX =$

$X^T X =$

Singular Vector Decomposition

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

WLOG, $n \geq d$

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

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

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

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

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

$XX =$

$X^T X =$

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

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 $r = \{ i : \sigma_i > 0 \}$
- Pseudo inverse $X^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$
- Condition number σ_1 / σ_d
- Matrix norms i.e., $\|X\|_2 = \sigma_1$, $\|X\|_F^2 = \sum_{i=1}^{d} \sigma_i^2$
- Principal Component Analysis

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

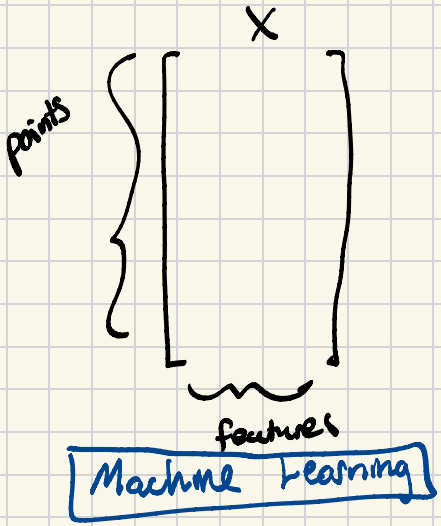
Thursday, March 26

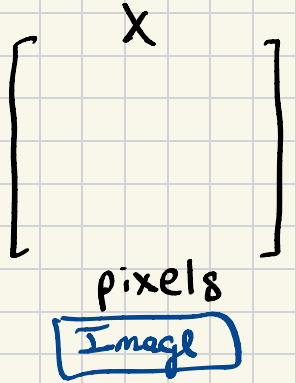
- Suyog Soti visits class on Tuesday!
- James Gnouen talk at noon in RN 15 on Thursday (no OH Thursday)

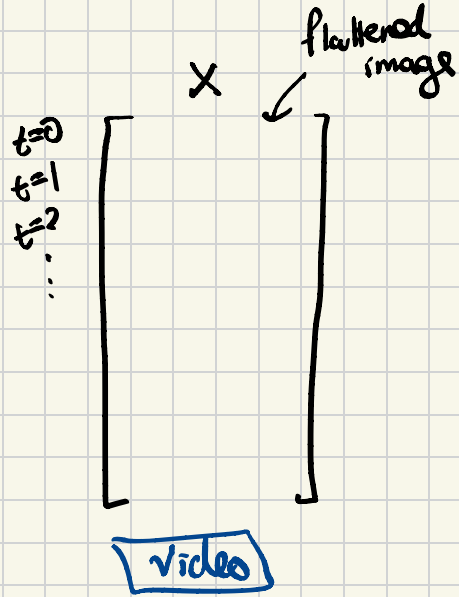
Today

- Best rank-k Approximation
- Compute with power method faster than $O(nd^2)$

Low Rank Approximation







Goal: Given X, best rank-k approximation

$$X_k = \text{argmin} \; \| X - XWW^T \|_F^2$$

XWW^T where $W \in \mathbb{R}^{d \times k}$ with orthonormal columns i.e. $W^TW = I$

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

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

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

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

$$= tr(X^TX) - tr(X^TXWW^T)$$

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

cyclic property of the trace

linearity of trace

$$\arg\min_W (*) = \arg\max_W tr(X^TXWW^T)$$

$$tr(X^TXWW^T) = tr(W^TX^TXW)$$

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

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

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

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

$z_i = \|v_i^T W\|_2^2$ $\quad 0 \le z_i \le 1$ 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 $z_i = \begin{cases} 1 & \text{if } i \le k \\ 0 & \text{else} \end{cases}$. $W = \sum_{i=1}^{k} v_i e_i^T$

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

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

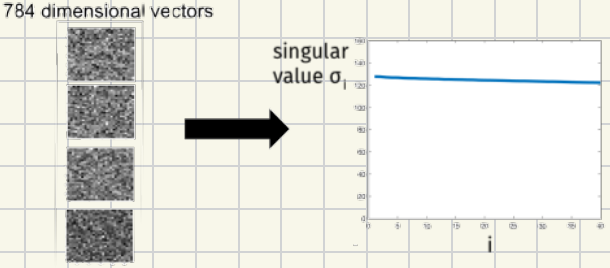
Q: How good is the approximation?

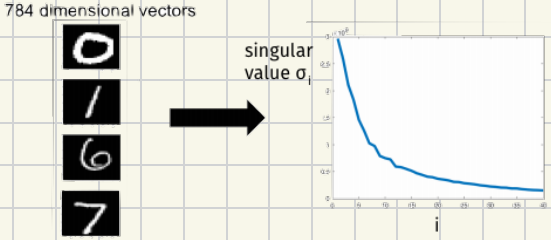
$$X - X_k = \sum_{i=1}^{d} \sigma_i u_i v_i^T - \sum_{i=1}^{k} \sigma_i u_i v_i^T = \sum_{i=k+1}^{d} \sigma_i u_i v_i^T$$

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

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

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





Power Method

Motivation: Compute X_k in $\tilde{O}(ndk)$ time instead of $O(nd^2)$ time

$$X^T X = \sum_{i=1}^{d} v_i v_i^T \sigma_i^2 = A = \sum_{i=1}^{d} v_i v_i^T \lambda_i$$

Goal: Compute z s.t. $\|z - v_i\|_2 \leq \epsilon$ in $O\left(nd \frac{\log d/\epsilon}{\gamma}\right)$ time

$z^{(0)} \sim \mathcal{N}(0, I_d)$

$z^{(0)} = c_1^{(0)} v_1 + c_2^{(0)} v_2 + \ldots + c_d^{(0)} v_d$ since $v_1, \ldots, v_d$ is a basis

$n^{(t)} = \|z^{(t)}\|_2$

$z^{(t+1)} = A z^{(t)} / n^{(t)}$

Return $z^{(q)}$ where $q = \frac{\log\left(d^{3.5}/\epsilon\right)}{\gamma}$

$$z^{(q)} = \frac{1}{\prod_{t} \pi_{\eta^{(t)}}} A^q z^{(0)}$$

Fact: $\left| \frac{c_j^{(0)}}{c_1^{(0)}} \right| \le d^3$ whp

Fact: $(1-\gamma)^{1/\gamma} \le 1/e$

$$\gamma = \frac{\lambda_1 - \lambda_2}{\lambda_1} \le \frac{\lambda_1 - \lambda_j}{\lambda_1}$$

$$\left| \frac{c_j^{(q)}}{c_1^{(q)}} \right| = \left| \frac{c_j^{(0)}}{c_1^{(0)}} \right| \left(\frac{\lambda_j}{\lambda_1} \right)^q \le d^3 \left(1 - \frac{\lambda_1 - \lambda_j}{\lambda_1} \right)^q \le d^3 \left((1-\gamma)^{1/\gamma} \right)^{\log\left(d^3 \sqrt{d/\epsilon}\right)} \le \sqrt{\epsilon/d}$$

Since $\sum_{i=1}^{d} (c_i^{(q)})^2 = 1 = (c_1^{(q)})^2 \left(1 + \sum_{i=2}^{d} \left| \frac{c_i^{(q)}}{c_1^{(q)}} \right|^2 \right) \le (c_1^{(q)})^2 (1 + \epsilon/2)$

$\Rightarrow |c_1^{(q)}| \gtrsim 1 - \epsilon$ (hiding some constants)

then $\| z^{(q)} - v_1 \|_2^2 = 2 - 2 \langle z^{(q)}, v_1 \rangle \overset{\text{WLOG}}{\le} 2 - 2(1-\epsilon) \le 2\epsilon$