TTIC 31150/CMSC 31150 Mathematical Toolkit (Fall 2024)

Avrim Blum

Lecture 7: SVD for Matrices

Recap

- Real Spectral Theorem (every self-adjoint operator has an orthonormal basis of eigenvectors, Raleigh quotients: $R_\varphi(v) = \langle \hat{v}, \varphi(\hat{v}) \rangle$, eigenvectors as maximizers/minimizers, positive semi-definiteness.
- Consider $\varphi: V \to W$. Analyze using eigenvectors/eigenvalues of $\varphi^*\varphi$ and $\varphi\varphi^*$.
- If v is eigenvector of $\varphi^*\varphi: V \to V$ with eigenvalue $\lambda \neq 0$, then $\varphi(v)$ is eigenvector of $\varphi\varphi^*: W \to W$ with eigenvalue λ; in other direction, $w, \varphi^*(w)$.
- If v_1, v_2 are orthogonal eigenvectors of $\varphi^*\varphi$ then $\varphi(v_1), \varphi(v_2)$ are orthogonal eigenvectors of $\varphi\varphi^*$.
- SVD: Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0$ be nonzero eigenvalues of $\varphi^*\varphi$ with corresponding orthonormal eigenvectors $v_1, \ldots, v_r$. Let $w_i = \varphi(v_i)/\sigma_i$. Then:
  - $w_1, \ldots, w_r$ are orthonormal, $\varphi(v_i) = \sigma_i w_i$ and $\varphi_i^*(w_i) = \sigma_i v_i$.
  - $\varphi = \sum_{i=1}^{r} \sigma_i \, |w_i\rangle\langle v_i|$, where $|w_i\rangle\langle v_i|$ is outer product.

SVD for Matrices

Let's consider matrices $A \in \mathbb{C}^{mxn}$, viewed as linear transformations from $\mathbb{C}^n$ to $\mathbb{C}^m$.

- Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0$ be nonzero singular values of A with $v_1, \ldots, v_r$ and $w_1, \ldots, w_r$ as the right and left singular vectors respectively.
  - $Av_i = \sigma_i w_i, A^* w_i = \sigma_i v_i$, where $A^* = \overline{A^T}$.
- Then,

$$A = \sum_{i=1}^{r} \sigma_i w_i v_i^* .$$

- Check: $(\sum_{i=1}^{r} \sigma_i w_i v_i^*) v_j = \sigma_j w_j v_j^* v_j = \sigma_j w_j = Av_j$, and if extend $v_1, \ldots, v_r$ to orthonormal basis, then for all other basis vectors both sides give 0.

SVD for Matrices

Let's consider matrices $A \in \mathbb{C}^{mxn}$, viewed as linear transformations from $\mathbb{C}^n$ to $\mathbb{C}^m$.

- Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0$ be nonzero singular values of A with $v_1, \ldots, v_r$ and $w_1, \ldots, w_r$ as the right and left singular vectors respectively.
  - $Av_i = \sigma_i w_i, A^* w_i = \sigma_i v_i$, where $A^* = \overline{A^T}$.
- Then,

$$A = \sum_{i=1}^{r} \sigma_i w_i v_i^* .$$

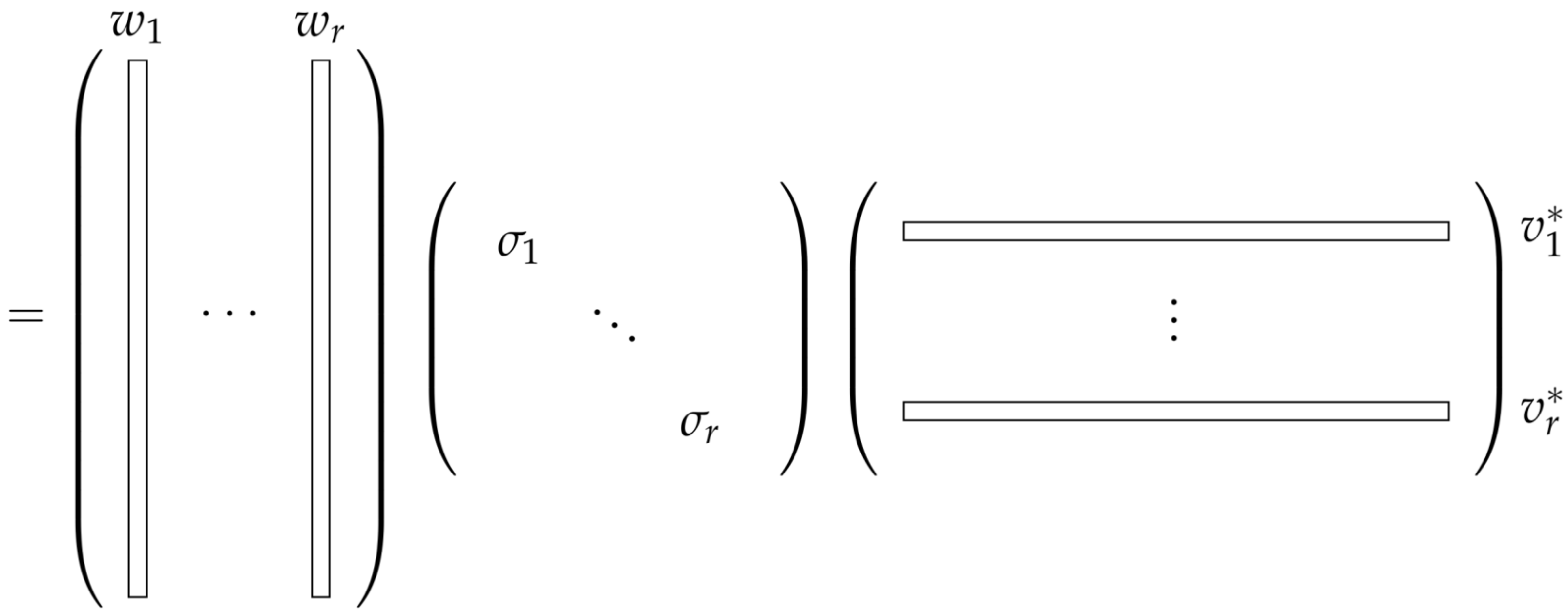
- Can write this as:

$$A = W\Sigma V^*$$

Where W has $w_1, \ldots, w_r$ as columns, V^* has $v_1^*, \ldots, v_r^*$ as rows, and Σ is an $r \times r$ diagonal matrix with $\Sigma_{ii} = \sigma_i$.

SVD for Matrices

$$A = W\Sigma V^*$$



SVD for Matrices

Definition 1.1 *A matrix $U \in \mathbb{C}^{n\times n}$ is known as a* unitary matrix *if the columns of U form an orthonormal basis for $\mathbb{C}^n$.*

If we complete w's and v's to an orthonormal bases, creating W_m and V_n respectively, these are unitary matrices.

Proposition 1.2 *Let $U \in \mathbb{C}^{n\times n}$ be a unitary matrix. Then $UU^* = U^*U = \mathrm{id}$, where* id *denotes the identity matrix.*

We had $A = W\Sigma V^*$. We can also write $A = W_m\Sigma' V_n^*$ where $\Sigma'_{ii} = \sigma_i$ for $i \leq r$ and all other entries of Σ' are 0.

SVD for Matrices

$$A = W_m \Sigma' V_n^*$$

$$= \begin{pmatrix} w_1 & \cdots & w_r & w_{r+1} & \cdots & w_m \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} v_1^* \\ \vdots \\ v_r^* \\ v_{r+1}^* \\ \vdots \\ v_n^* \end{pmatrix}$$

$Av_i = \sigma_i w_i$

$AV = (W\Sigma' V^*)V = W\Sigma'$

$A^*AV =$

$AA^*W =$

Low-rank approximation for matrices

Given matrix A, we may want to find the matrix B of rank $\leq k$ that "best approximates" A.

What notion of approximation?

We'll use spectral norm: $$\|(A-B)\|_2 = \max_{v \neq 0} \frac{\|(A-B)v\|_2}{\|v\|_2}.$$

For $v = (c_1, \cdots, c_n)^\top$,

$$\|v\|_2 = \langle v, v \rangle = \left(\sum_{i=1}^n |c_i|^2\right)^{1/2}$$

Next class will see also works for Frobenius norm $= \sqrt{\sum_{ij} (A-B)_{ij}{}^2}$.

Solution: take top k singular vectors: $B = A_k = \sum_{i=1}^k \sigma_i w_i v_i^*$.

Low-rank approximation for matrices

Proposition 2.1 $\|A - A_k\|_2 = \sigma_{k+1}.$

$$\|(A - B)\|_2 = \max_{v \neq 0} \frac{\|(A - B)v\|_2}{\|v\|_2}.$$

Let's start with the easier "≥" direction:

What v should we try?

$$(A - A_k)v_{k+1} = \left(\sum_{i=k+1}^{r} \sigma_i w_i v_i^*\right) v_{k+1} = \sigma_{k+1} w_{k+1}.$$

Length is σ_{k+1}.

Low-rank approximation for matrices

Proposition 2.1 $\|A - A_k\|_2 \;=\; \sigma_{k+1}.$

$$\|(A-B)\|_2 \;=\; \max_{v\neq 0} \frac{\|(A-B)v\|_2}{\|v\|_2}.$$

Now let's do the "≤" direction:

Write v as linear combination of $v_1, \ldots, v_r$ plus orthogonal component. Orthogonal part in nullspace.

- $(A - A_k)v = \left(\sum_{i=k+1}^{r} \sigma_i w_i v_i^*\right)\left(\sum_{i=1}^{r} c_i v_i\right) \;=\; \sum_{i=k+1}^{r} c_i \sigma_i w_i$
- $\|(A - A_k)v\|^2 = \|\sum_{i=k+1}^{r} c_i \sigma_i w_i\|^2 = \sum_{i=k+1}^{r} |c_i|^2 |\sigma_i|^2$
- We can wlog assume $\|v\| = 1$. What does this say about $\sum_{i=k+1}^{r} |c_i|^2$? Ans: ≤ 1.
- So, $\sum_{i=k+1}^{r} |c_i|^2 |\sigma_i|^2$ is maximized at $c_{k+1} = 1$. Get $\|(A - A_k)v\|_2^2 \leq \sigma_{k+1}^2$.

Low-rank approximation for matrices

Proposition 2.1 $\|A - A_k\|_2 = \sigma_{k+1}.$

$$\|(A-B)\|_2 = \max_{v \neq 0} \frac{\|(A-B)v\|_2}{\|v\|_2}.$$

Now, just need to show that no other rank-k approximation can get closer.

But first, note that our argument also shows that $\|A\|_2 = \sigma_1$.

- $Av_1 = \left(\sum_{i=1}^{r} \sigma_i w_i v_i^*\right) v_1 = \sigma_1 w_1$. Length is σ_1.
- $Av = \left(\sum_{i=1}^{r} \sigma_i w_i v_i^*\right)\left(\sum_{i=1}^{r} c_i v_i\right) = \sum_{i=1}^{r} c_i \sigma_i w_i$. $\|Av\|^2 = \sum_{i=1}^{r} c_i^2 \sigma_i^2 \leq \sigma_1^2$.

Low-rank approximation for matrices

$$\|(A-B)\|_2 = \max_{v \neq 0} \frac{\|(A-B)v\|_2}{\|v\|_2}.$$

Proposition 2.4 *Let $B \in \mathbb{C}^{m \times n}$ have* $\text{rank}(B) \leq k$ *and let $k < r$. Then $\|A - B\|_2 \geq \sigma_{k+1}$.*

Proof: (very similar to proof for Courant-Fischer thm)

- Since $rank(B) \leq k$, the nullspace of B has dimension $\geq n - k$. (rank-nullity thm)
- So, (nullspace of B) ∩ span$(v_1, \ldots, v_{k+1})$ is a subspace of dimension ≥ 1. Pick some unit-length $\hat{z} = \sum_{1 \leq i \leq k+1} c_i v_i$ in this intersection.
- We have $(A - B)\hat{z} = A\hat{z} - B\hat{z} = A\hat{z}$, so:
- $\|(A-B)\hat{z}\|_2^2 = \|A\hat{z}\|_2^2 = \langle A\hat{z}, A\hat{z}\rangle = \langle \sum_{1 \leq i \leq k+1} c_i \sigma_i w_i, \sum_{1 \leq i \leq k+1} c_i \sigma_i w_i \rangle$
  $= \sum_{1 \leq i \leq k+1} |c_i|^2 \sigma_i^2 \geq (\sum_{1 \leq i \leq k+1} |c_i|^2)\sigma_{k+1}^2 = \sigma_{k+1}^2$

Midterm next Monday

- In class (TTIC 530)
- You may bring in one sheet of notes.