TTIC 31150/CMSC 31150 Mathematical Toolkit (Fall 2024)

Avrim Blum

Lecture 6: SVD

Recap

- Any finite-dimensional inner product space has orthonormal basis. Fourier coefficients, Parseval's identity. Adjoint of linear transform. Reisz representation theorem. Self-adjoint linear operators: eigenvalues are real and eigenvectors corresponding to distinct eigenvalues are orthogonal.

- Real Spectral Theorem: every self-adjoint operator $\varphi: V \to V$ for finite-dimensional V has an orthonormal basis of eigenvectors (i.e., is "orthogonally diagonalizable").

- Raleigh quotients: $R_\varphi(v) = \langle \hat{v}, \varphi(\hat{v}) \rangle$ where $\hat{v} = v/\|v\|$

- The vector v such that applying φ gives the largest "stretch" in $\hat{v}$ direction is the eigenvector of largest eigenvalue, and likewise for the evector of smallest evalue. (Extension: Courant-Fischer Theorem)

- Positive semidefiniteness (see next slide).

Positive Semidefiniteness (recap)

Definition 3.4 *Let $\varphi : V \to V$ be a self-adjoint operator. φ is said to be* positive semidefinite *if $\mathcal{R}_\varphi(v) \geq 0$ for all $v \neq 0$. φ is said to be* positive definite *if $\mathcal{R}_\varphi(v) > 0$ for all $v \neq 0$.*

Proposition 3.5 *Let $\varphi : V \to V$ be a self-adjoint linear operator. Then the following are equivalent:*

1. *$\mathcal{R}_\varphi(v) \geq 0$ for all $v \neq 0$.*
2. *All eigenvalues of φ are non-negative.*
3. *There exists $\alpha : V \to V$ such that $\varphi = \alpha^*\alpha$.*

Part of argument: if $\varphi = \alpha^*\alpha$ then $\langle v, \varphi(v)\rangle = \langle v, \alpha^*(\alpha(v))\rangle = \langle \alpha(v), \alpha(v)\rangle \geq 0$. This also means that if v is an eigenvector, its eigenvalue must be non-negative.

The decomposition of a positive semidefinite operator in the form $\varphi = \alpha^*\alpha$ is known as the Cholesky decomposition of the operator. Note that if we can write φ as $\alpha^*\alpha$ for any $\alpha : V \to W$, then this in fact also shows that φ is self-adjoint and positive semidefinite.

Singular Value Decomposition preliminaries

- Consider a linear transformation $\varphi: V \to W$. We can use our previous discussion to analyze the eigenvectors of $\varphi^*\varphi: V \to V$ and $\varphi\varphi^*: W \to W$, and then use these to get a nice decomposition of φ called **Singular Value Decomposition (SVD).**

Proposition 1.1 *Let $\varphi : V \to W$ be a linear transformation. Then $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$ are self-adjoint positive semidefinite linear operators with the same non-zero eigenvalues.*

Self-adjointness of $\varphi\varphi^*$ (the proof for $\varphi^*\varphi$ is analogous):

- $\langle w_1, \varphi(\varphi^*(w_2))\rangle = \langle \varphi^*(w_1), \varphi^*(w_2)\rangle = \langle \varphi(\varphi^*(w_1)), w_2\rangle$.

Using fact that $\varphi^{**} = \varphi$ (see hwk2 q2)

Positive semidefiniteness of $\varphi\varphi^*$ (the proof for $\varphi^*\varphi$ is analogous):

- $\langle w, \varphi(\varphi^*(w))\rangle = \langle \varphi^*(w), \varphi^*(w)\rangle \geq 0$.

Singular Value Decomposition preliminaries

- Consider a linear transformation $\varphi: V \to W$. We can use our previous discussion to analyze the eigenvectors of $\varphi^*\varphi: V \to V$ and $\varphi\varphi^*: W \to W$, and then use these to get a nice decomposition of φ called **Singular Value Decomposition (SVD).**

Proposition 1.1 *Let $\varphi : V \to W$ be a linear transformation. Then $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$ are self-adjoint positive semidefinite linear operators with the same non-zero eigenvalues.*

Now just need to show they have the same nonzero eigenvalues:

- Let $\lambda > 0$ be an eigenvalue of $\varphi^*\varphi$ with eigenvector v. So $\varphi^*\big(\varphi(v)\big) = \lambda v$.
- This implies $\varphi\Big(\varphi^*\big(\varphi(v)\big)\Big) = \lambda\varphi(v)$. Note that $\varphi(v)$ can't be 0 (by ↑), so $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ of eigenvalue λ.

Singular Value Decomposition preliminaries

- Consider a linear transformation $\varphi: V \to W$. We can use our previous discussion to analyze the eigenvectors of $\varphi^*\varphi: V \to V$ and $\varphi\varphi^*: W \to W$, and then use these to get a nice decomposition of φ called **Singular Value Decomposition (SVD).**

Proposition 1.1 *Let $\varphi : V \to W$ be a linear transformation. Then $\varphi^*\varphi : V \to V$ and $\varphi\varphi^* : W \to W$ are self-adjoint positive semidefinite linear operators with the same non-zero eigenvalues.*

- This implies $\varphi\left(\varphi^*(\varphi(v))\right) = \lambda\varphi(v)$. Note that $\varphi(v)$ can't be 0 (by ↑), so $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ of eigenvalue λ.

Proposition 1.2 *Let v be an eigenvector of $\varphi^*\varphi$ with eigenvalue $\lambda \neq 0$. Then $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ with eigenvalue λ. Similarly, if w is an eigenvector of $\varphi\varphi^*$ with eigenvalue $\lambda \neq 0$, then $\varphi^*(w)$ is an eigenvector of $\varphi^*\varphi$ with eigenvalue λ.*

Singular Value Decomposition preliminaries

Proposition 1.2 *Let v be an eigenvector of $\varphi^*\varphi$ with eigenvalue $\lambda \neq 0$. Then $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ with eigenvalue λ. Similarly, if w is an eigenvector of $\varphi\varphi^*$ with eigenvalue $\lambda \neq 0$, then $\varphi^*(w)$ is an eigenvector of $\varphi^*\varphi$ with eigenvalue λ.*

Proposition 1.3 *Let the subspaces V_λ and W_λ be defined as*

$$V_\lambda \;:=\; \{v \in V \mid \varphi^*\varphi(v) = \lambda \cdot v\} \;\; \textit{and} \;\; W_\lambda \;:=\; \{w \in W \mid \varphi\varphi^*(w) = \lambda \cdot w\} \; .$$

Then for any $\lambda \neq 0$, $\dim(V_\lambda) = \dim(W_\lambda)$.

Proof:

- If $\dim(V_\lambda) = k$ then we have k orthogonal eigenvectors $v_1, \ldots, v_k$ of $\varphi^*\varphi$ with eigenvalue λ. So, $\varphi(v_1), \ldots \varphi(v_k)$ are eigenvectors of $\varphi\varphi^*$ with eigenvalue λ. In fact, they're also orthogonal: $\langle \varphi(v_i), \varphi(v_j) \rangle = \langle \varphi^*\varphi(v_i), v_j \rangle = \langle \lambda v_i, v_j \rangle = 0$. So, $\dim(W_\lambda) \geq k$. And vice versa.

Singular Value Decomposition preliminaries

Proposition 1.2 *Let v be an eigenvector of $\varphi^*\varphi$ with eigenvalue $\lambda \neq 0$. Then $\varphi(v)$ is an eigenvector of $\varphi\varphi^*$ with eigenvalue λ. Similarly, if w is an eigenvector of $\varphi\varphi^*$ with eigenvalue $\lambda \neq 0$, then $\varphi^*(w)$ is an eigenvector of $\varphi^*\varphi$ with eigenvalue λ.*

Proposition 1.3 *Let the subspaces V_λ and W_λ be defined as*

$$V_\lambda := \{v \in V \mid \varphi^*\varphi(v) = \lambda \cdot v\} \;\; \textit{and} \;\; W_\lambda := \{w \in W \mid \varphi\varphi^*(w) = \lambda \cdot w\} .$$

Then for any $\lambda \neq 0$, $\dim(V_\lambda) = \dim(W_\lambda)$.

Using this, we now get…

Singular Value Decomposition

Proposition 1.4 *Let* $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ *be the non-zero eigenvalues of* $\varphi^*\varphi$, *and let* $v_1, \ldots, v_r$ *be a corresponding orthonormal eigenvectors (since* $\varphi^*\varphi$ *is self-adjoint, these are a subset of some orthonormal eigenbasis). For* $w_1, \ldots, w_r$ *defined as* $w_i = \varphi(v_i)/\sigma_i$, *we have that*

1. $\{w_1, \ldots, w_r\}$ *form an orthonormal set.*
2. *For all* $i \in [r]$
$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) = \sigma_i \cdot v_i \,.$$

So, even though φ and φ^* don't have eigenvectors (their domain and range are different – they are arbitrary linear transformations / matrices), the v_i and w_i are a bit like eigenvectors. They are called the (right and left) *singular vectors*, and the σ_i are called *singular values*.

Singular Value Decomposition

Proposition 1.4 *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ be the non-zero eigenvalues of $\varphi^*\varphi$, and let $v_1, \ldots, v_r$ be a corresponding orthonormal eigenvectors (since $\varphi^*\varphi$ is self-adjoint, these are a subset of some orthonormal eigenbasis). For $w_1, \ldots, w_r$ defined as $w_i = \varphi(v_i)/\sigma_i$, we have that*

1. *$\{w_1, \ldots, w_r\}$ form an orthonormal set.*

2. *For all $i \in [r]$*

$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) = \sigma_i \cdot v_i .$$

Proof of (1):

- We already saw orthogonal. Unit length because $\langle \varphi(v_i), \varphi(v_i) \rangle = \langle \varphi^*\varphi(v_i), v_i \rangle = \sigma_i^2$.

Singular Value Decomposition

Proposition 1.4 *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ be the non-zero eigenvalues of $\varphi^* \varphi$, and let $v_1, \ldots, v_r$ be a corresponding orthonormal eigenvectors (since $\varphi^* \varphi$ is self-adjoint, these are a subset of some orthonormal eigenbasis). For $w_1, \ldots, w_r$ defined as $w_i = \varphi(v_i) / \sigma_i$, we have that*

1. *$\{w_1, \ldots, w_r\}$ form an orthonormal set.*
2. *For all $i \in [r]$*

$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) = \sigma_i \cdot v_i .$$

Proof of (2):

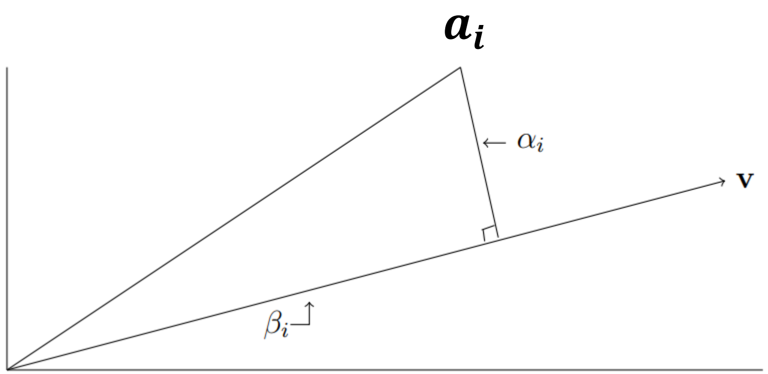
- $\varphi(v_i) = \sigma_i w_i$ by definition.
- $\varphi^*(w_i) = \varphi^*(\varphi(v_i)/\sigma_i) = \sigma_i^2 v_i/\sigma_i = \sigma_i v_i$.

Singular Value Decomposition

Proposition 1.4 *Let $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ be the non-zero eigenvalues of $\varphi^*\varphi$, and let $v_1, \ldots, v_r$ be a corresponding orthonormal eigenvectors (since $\varphi^*\varphi$ is self-adjoint, these are a subset of some orthonormal eigenbasis). For $w_1, \ldots, w_r$ defined as $w_i = \varphi(v_i)/\sigma_i$, we have that*

1. *$\{w_1, \ldots, w_r\}$ form an orthonormal set.*
2. *For all $i \in [r]$*

$$\varphi(v_i) = \sigma_i \cdot w_i \quad \text{and} \quad \varphi^*(w_i) =$$



Matrix view: $Av_i = \sigma_i w_i$ and $A^T w_i = \sigma_i v_i$.

- If you view the rows of A as representing m points in n-dimensional space, then $span(v_1, \ldots, v_k)$ will be the "best-fitting" k-dimensional subspace in the sense of minimizing the sum of squared distances to the subspace.
  - Minimizing squared distance is equivalent to maximizing squared projection
  - The quantity $\hat{v}^T A^T A \hat{v} = \langle \hat{v}, A^T A \hat{v} \rangle$ is the sum of projections along v squared, maximized at v_1 because this is the Raleigh quotient.

Singular Value Decomposition

Definition 1.6 *Let V, W be inner product spaces and let $v \in V, w \in W$ be any two vectors. The* outer product *of w with v, denoted as $|w\rangle\langle v|$, is a linear transformation from V to W such that*

$$|w\rangle\langle v|(u) := \langle v, u\rangle \cdot w .$$

Matrix view: This is the rank-1 matrix wv^T (as opposed to the inner product w^Tv).

- Get $wv^Tu = w(v^Tu)$.

Why is wv^T rank 1?

- Because all rows are multiples of v^T (and all columns are multiples of w).

We now get…

Singular Value Decomposition

Proposition 1.8 *Let V, W be finite dimensional inner product spaces and let $\varphi : V \to W$ be a linear transformation with non-zero singular values $\sigma_1, \ldots, \sigma_r$, right singular vectors $v_1, \ldots, v_r$ and left singular vectors $w_1, \ldots, w_r$. Then,*

$$\varphi = \sum_{i=1}^{r} \sigma_i \cdot |w_i\rangle \langle v_i| \, .$$

$$A = \sum_{i=1}^{r} \sigma_i \, w_i v_i^T$$

This is the Singular Value Decomposition of φ (or A).

Singular Value Decomposition

Proposition 1.8 *Let V, W be finite dimensional inner product spaces and let $\varphi : V \to W$ be a linear transformation with non-zero singular values $\sigma_1, \ldots, \sigma_r$, right singular vectors $v_1, \ldots, v_r$ and left singular vectors $w_1, \ldots, w_r$. Then,*

$$\varphi = \sum_{i=1}^{r} \sigma_i \cdot |w_i\rangle \langle v_i| .$$

$$A = \sum_{i=1}^{r} \sigma_i \, w_i v_i^T$$

Proof:

- First, note that the RHS is a linear transformation, so we just need to show it acts correctly on basis vectors.
- Let's define a basis: take $v_1, \ldots, v_r$ and extend arbitrarily to orthonormal basis.
- What is RHS applied to v_j? Ans: $\sigma_j w_j = \varphi(v_j)$.
- All the rest of the basis vectors are in the null-space. LHS and RHS both evaluate to 0.