Lecture 2 - Linear Algebra

Konstantinos Ameranis

Thursday 28th September, 2023

THE UNIVERSITY OF CHICAGO

Overview

Linear Algebra review

Vector-vector inner product

$$\mathbf{x}^T\mathbf{y} \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Vector-vector outer product

$$\mathbf{x}\mathbf{y}^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}$$

Linear Algebra Review

Matrix-vector product

$$
\mathbf{Ax} \in \mathbb{R}^m = \begin{bmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_m & — \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T \mathbf{x} \\ \mathbf{a}_2^T \mathbf{x} \\ \vdots \\ \mathbf{a}_m^T \mathbf{x} \end{bmatrix}
$$

Matrix-matrix product

$$
\mathbf{AB} \in \mathbb{R}^{m \times n} = \begin{bmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_m & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \\ | & | & & | \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{a}_1^T \mathbf{b}_1 & \mathbf{a}_1^T \mathbf{b}_2 & \cdots & \mathbf{a}_1^T \mathbf{b}_n \\ \mathbf{a}_2^T \mathbf{b}_2 & \mathbf{a}_2^T \mathbf{b}_2 & \cdots & \mathbf{a}_2^T \mathbf{b}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m^T \mathbf{b}_1 & \mathbf{a}_m^T \mathbf{b}_2 & \cdots & \mathbf{a}_m^T \mathbf{b}_n \end{bmatrix}
$$

Vector Norms

A norm measures in some way the size or magnitude of a vector or matrix.
Properties:

- Triangle inequality $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- Absolute homogeneity $\forall \alpha \in \mathbb{R} \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
- Positive definiteness $\forall \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

p-norms: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ Examples:

- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- $\|\mathbf{x}\|_\infty = \max_i |x_i|$

If p is skipped, it's implied that $p = 2$.
Given a matrix $0 \prec \mathbf{H} \in \mathbb{R}^{n \times n}, \|\mathbf{x}\|_{\mathbf{H}} = \sqrt{\mathbf{x}^T \mathbf{H} \mathbf{x}}$

Matrix Norms

Similarly to vector norms, matrix norms have the same properties. The Euclidean norm $\|\cdot\|_F$ for matrices is called the Frobenius norm

- $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} A_{ij}^2}$
- $\|\mathbf{A}\|_{\alpha,\beta} = \max_{\mathbf{x}} \frac{\|\mathbf{Ax}\|_\beta}{\|\mathbf{x}\|_\alpha}$
- $\|\mathbf{A}\|_{2,2} = \|\mathbf{A}\|_2 = \sqrt{\sigma_{\max}(\mathbf{A})}$
- $\|\mathbf{A}\|_{1,\infty} = \|\mathbf{A}\|_\infty = \max_{ij} |A_{ij}|$

Eigenvectors and eigenvalues

In this course we are going to deal primarily with real, square matrices. All such matrices have solutions to the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Those solutions $\mathbf{x}$ are called the **eigenvectors** of $\mathbf{A}$ and λ is the corresponding **eigenvalue**.

Arranging the eigenvectors in matrix $\mathbf{V}$and the eigenvalues in the diagonal matrix $\mathbf{\Lambda}$ you get the **eigenvector decomposition** of matrix $\mathbf{A}$.

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

Traditionally, the eigenvalues are ordered from largest to smallest $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. This sequence is known as the **spectrum** of matrix $\mathbf{A}$.

This tradition is based mostly on PCA where the top k eigenvalues are kept. In this presentation I am following the **opposite** convention because we usually care about the **smallest** eigenvectors of the Laplacian matrix.

The Real Spectral Theorem

Theorem (Real Spectral Theorem)

If $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is a symmetric matrix, then there exists an orthonormal basis* $V = \begin{bmatrix} — & \mathbf{v}_1 & — \\ — & \mathbf{v}_2 & — \\ & \vdots & \\ — & \mathbf{v}_m & — \end{bmatrix}$ *consisting of eigenvectors of* $\mathbf{A}$ *and each corresponding eigenvalue is real.*

There exists *some* eigenvalue

Lemma

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ *be a symmetric matrix. Then* $\mathbf{A}$ *has at least one eigenvalue.*

Take the Rayleigh quotient of $\mathbf{x}$ with respect to matrix $\mathbf{A}$

$$f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

Since the Rayleigh Quotient is homogeneous $f(\alpha \cdot \mathbf{x}) = \frac{(\alpha \cdot \mathbf{x})^T \mathbf{A} (\alpha \cdot \mathbf{x})}{(\alpha \cdot \mathbf{x})^T (\alpha \cdot \mathbf{x})}$ $= \frac{\alpha^2}{\alpha^2} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = f(\mathbf{x})$, it suffices to consider unit vectors of $\mathbf{x}$. The radius 1 ℓ_2-ball is a closed and compact set and therefore $f(\mathbf{x})$ achieves its maximum at one point in that set.

There exists *some* eigenvalue

Let $\mathbf{x}$ be a non-zero vector that maximizes $f(\mathbf{x})$. According to first order optimality conditions, the gradient of $f(\mathbf{x})$ must be zero.

$$\nabla f(\mathbf{x}) = \nabla \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{(\mathbf{x}^T \mathbf{x})(2\mathbf{A}\mathbf{x}) - (\mathbf{x}^T \mathbf{A} \mathbf{x})(2\mathbf{x})}{(\mathbf{x}^T \mathbf{x})^2} = 0$$

$$(\mathbf{x}^T \mathbf{x})(\mathbf{A}\mathbf{x}) = (\mathbf{x}^T \mathbf{A} \mathbf{x})\mathbf{x} \Leftrightarrow \mathbf{A}\mathbf{x} = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \mathbf{x}$$

This proves that the maximizer $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ and its Rayleigh quotient is its corresponding eigenvalue. As $\mathbf{x}$ maximizes the Rayleigh quotient, this eigenvalue must be λ_n.

Eigenvectors are orthonormal

Using induction we can prove that eigenvectors are orthonormal. The base case is covered above. Assuming that $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$ are orthonormal, we can prove that $\mathbf{v}_k$ is orthonormal to $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$.

$$\mathbf{A}_k = M - \sum_{i=k+1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

. For $n \geq j \geq k+1$ we have

$$\mathbf{A}_k \mathbf{v}_j = \mathbf{A}\mathbf{v}_j - \sum_{i=k+1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \mathbf{v}_j = 0$$

For vectors orthogonal to $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$
$\mathbf{A}_k \mathbf{x} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{x}^T \mathbf{A}_k \mathbf{x} = \mathbf{x}^T \mathbf{A}\mathbf{x}$

Eigenvectors are orthonormal

- Let $\mathbf{y}$ to be the unit maximizer of $\mathbf{y}^T \mathbf{A}_k \mathbf{y}$[1].

[1] In order to get a non-zero eigenvalue as the maximum one, the matrix needs to be positive semi-definite. The matrix $\tilde{\mathbf{A}} = \mathbf{A} + (1 - \lambda_1)\mathbf{I} \succ 1$ is PSD and has the same eigenvectors as $\mathbf{A}$

Eigenvectors are orthonormal

- Let $\mathbf{y}$ to be the unit maximizer of $\mathbf{y}^T \mathbf{A}_k \mathbf{y}$[1].
- Similarly to above, $\mathbf{y}$ is an eigenvector of $\mathbf{A}_k$.

[1] In order to get a non-zero eigenvalue as the maximum one, the matrix needs to be positive semi-definite. The matrix $\tilde{\mathbf{A}} = \mathbf{A} + (1 - \lambda_1)\mathbf{I} \succ 1$ is PSD and has the same eigenvectors as $\mathbf{A}$

Eigenvectors are orthonormal

- Let $\mathbf{y}$ to be the unit maximizer of $\mathbf{y}^T\mathbf{A}_k\mathbf{y}$[1].
- Similarly to above, $\mathbf{y}$ is an eigenvector of $\mathbf{A}_k$.
- Let $\tilde{\mathbf{y}} = \mathbf{y} - \sum_{i=k+1}^{n} \mathbf{v}_i(\mathbf{v}_i\mathbf{y})$ be the orthogonal projection to $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$ and let $\hat{\mathbf{y}} = \tilde{\mathbf{y}}/\|\tilde{\mathbf{y}}\|$. We know that $\tilde{\mathbf{y}}^T\mathbf{A}_k\tilde{\mathbf{y}} = \mathbf{y}^T\mathbf{A}\mathbf{y}$.

[1]In order to get a non-zero eigenvalue as the maximum one, the matrix needs to be positive semi-definite. The matrix $\tilde{\mathbf{A}} = \mathbf{A} + (1-\lambda_1)\mathbf{I} \succ 1$ is PSD and has the same eigenvectors as $\mathbf{A}$

Eigenvectors are orthonormal

- Let $\mathbf{y}$ to be the unit maximizer of $\mathbf{y}^T\mathbf{A}_k\mathbf{y}$[1].
- Similarly to above, $\mathbf{y}$ is an eigenvector of $\mathbf{A}_k$.
- Let $\tilde{\mathbf{y}} = \mathbf{y} - \sum_{i=k+1}^{n} \mathbf{v}_i(\mathbf{v}_i\mathbf{y})$ be the orthogonal projection to $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$ and let $\hat{\mathbf{y}} = \tilde{\mathbf{y}}/\|\tilde{\mathbf{y}}\|$. We know that $\tilde{\mathbf{y}}^T\mathbf{A}_k\tilde{\mathbf{y}} = \mathbf{y}^T\mathbf{A}\mathbf{y}$.
- If $\mathbf{y}$ is not orthogonal to the previous eigenvectors then $\|\tilde{\mathbf{y}}\| < \|\mathbf{y}\|$ and since $\tilde{\mathbf{y}}^T\mathbf{A}\tilde{\mathbf{y}} > 0$, then $\hat{\mathbf{y}}^T\mathbf{A}\hat{vy} > \mathbf{y}\mathbf{A}_k\mathbf{y}$, which is a contradiction.

[1] In order to get a non-zero eigenvalue as the maximum one, the matrix needs to be positive semi-definite. The matrix $\tilde{\mathbf{A}} = \mathbf{A} + (1 - \lambda_1)\mathbf{I} \succ 1$ is PSD and has the same eigenvectors as $\mathbf{A}$

Eigenvectors are orthonormal

- Let $\mathbf{y}$ to be the unit maximizer of $\mathbf{y}^T \mathbf{A}_k \mathbf{y}$[1].
- Similarly to above, $\mathbf{y}$ is an eigenvector of $\mathbf{A}_k$.
- Let $\tilde{\mathbf{y}} = \mathbf{y} - \sum_{i=k+1}^{n} \mathbf{v}_i(\mathbf{v}_i\mathbf{y})$ be the orthogonal projection to $\mathbf{v}_n, \cdots, \mathbf{v}_{k+1}$ and let $\hat{\mathbf{y}} = \tilde{\mathbf{y}}/\|\tilde{\mathbf{y}}\|$. We know that $\tilde{\mathbf{y}}^T \mathbf{A}_k \tilde{\mathbf{y}} = \mathbf{y}^T \mathbf{A} \mathbf{y}$.
- If $\mathbf{y}$ is not orthogonal to the previous eigenvectors then $\|\tilde{\mathbf{y}}\| < \|\mathbf{y}\|$ and since $\tilde{\mathbf{y}}^T \mathbf{A} \tilde{\mathbf{y}} > 0$, then $\hat{\mathbf{y}}^T \mathbf{A} \hat{vy} > \mathbf{y} \mathbf{A}_k \mathbf{y}$, which is a contradiction.
- Therefore $\mathbf{y}$ is the k-th eigenvector of $\mathbf{A}$ and orthonormal to all larger eigenvectors. This concludes the inductive proof.

[1] In order to get a non-zero eigenvalue as the maximum one, the matrix needs to be positive semi-definite. The matrix $\tilde{\mathbf{A}} = \mathbf{A} + (1 - \lambda_1)\mathbf{I} \succ 1$ is PSD and has the same eigenvectors as $\mathbf{A}$

Courant-Fisher Theorem

Theorem

Courant-Fisher Theorem Let $\mathbf{A}$ *be a symmetric matrix with eigenvalues* $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. *Then,*

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ dim(S)=k}} \max_{\substack{\mathbf{x} \in S \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\substack{T \subseteq \mathbb{R}^n \\ dim(T)=n-k+1}} \min_{\substack{\mathbf{x} \in T \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

where the maximization and minimization are over subspaces S *and* T *of* $\mathbb{R}^n$

Courant-Fisher Theorem

We will prove one side. The same proof also works for the other side.

Proof.

Consider $U = \text{span}(\{\mathbf{v}_k, \mathbf{v}_{k+1}, \cdots, \mathbf{v}_n\})$.
You can see that $\dim(S) + \dim(U) = n + 1$, which means that $\dim(S \cap U) \geq 1$.
From the above it is shown that $\exists \mathbf{x}1 = \sum_{i=k}^{n} c_i \mathbf{v}_i \in S$ with

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum_{i=k}^{n} c_i^2} \leq \frac{\sum_{i=k}^{n} \lambda_k c_i^2}{\sum_{i=k}^{n} c_i^2} = \lambda_k$$

The vector $\mathbf{x} = \mathbf{v}_k$ shows this inequality to be tight. $\square$