Lecture 3: Linear algebra (part 2)

STATS 101: Foundations of Statistics

Linh Tran
ThetaHat.AI@gmail.com

December 5, 2019

Announcements

- Review survey
- Question: How long did the homework take?
- Next assignment will be posted tonight (due 12/11 @ 11:59pm)
- Classes will be recorded going forward

Outline

All things linear algebra

- Operations and Properties
- Eigenvalues & eigenvectors
- Matrix decomposition
- Matrix Calculus

Norms

A *norm* of a vector $\mathbf{x}$, denoted $||\mathbf{x}||$ is a measure of the "*length*" of the vector. For example, the ℓ_2-norm (aka Euclidean norm) is

$$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1}$$

n.b. $||\mathbf{x}||_2^2 = \mathbf{x}^\top \mathbf{x}$, i.e. the squared norm of a vector is the dot product with itself.

Norms

A *norm* of a vector $\mathbf{x}$, denoted $||\mathbf{x}||$ is a measure of the "*length*" of the vector. For example, the ℓ_2-norm (aka Euclidean norm) is

$$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1}$$

n.b. $||\mathbf{x}||_2^2 = \mathbf{x}^\top \mathbf{x}$, i.e. the squared norm of a vector is the dot product with itself.

Other norms:

- ℓ_1-norm, i.e. $||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$.
- ℓ_∞-norm, i.e. $||\mathbf{x}||_\infty = \max_i |x_i|$.
- ℓ_p-norm, i.e. $||\mathbf{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$.

Norms

Formally, a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying four properties:

1. $\forall \mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) \geq 0$ (non-negativity).
2. $f(\mathbf{x}) = 0$ iff $\mathbf{x} = 0$ (definiteness).
3. $\forall \mathbf{x} \in \mathbb{R}^n, c \in \mathbb{R}, f(c\mathbf{x}) = |c| f(\mathbf{x})$ (homogeneity).
4. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (triangle inequality).

Norms

Formally, a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying four properties:

1. $\forall \mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) \geq 0$ (non-negativity).
2. $f(\mathbf{x}) = 0$ iff $\mathbf{x} = 0$ (definiteness).
3. $\forall \mathbf{x} \in \mathbb{R}^n, c \in \mathbb{R}, f(c\mathbf{x}) = |c| f(\mathbf{x})$ (homogeneity).
4. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (triangle inequality).

Norms can also be defined for matrices, e.g. The Frobenius norm,

$$||\mathbf{A}||^F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{A}_{ij}^2} = \sqrt{tr(\mathbf{A}^\top \mathbf{A})} \tag{2}$$

Linear independence

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \in \mathbb{R}^m$ is *(linearly) dependent* if one of the vectors $\mathbf{x}_i$ can be represented as a linear combination of the remaining vectors, i.e.

$$\mathbf{x}_n = \sum_{i=1}^{n-1} \alpha_i \mathbf{x}_i \tag{3}$$

for some scalar values $\alpha_1, \alpha_2, \ldots, \alpha_{n-1} \in \mathbb{R}$

Linear independence

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \in \mathbb{R}^m$ is *(linearly) dependent* if one of the vectors $\mathbf{x}_i$ can be represented as a linear combination of the remaining vectors, i.e.

$$\mathbf{x}_n = \sum_{i=1}^{n-1} \alpha_i \mathbf{x}_i \tag{3}$$

for some scalar values $\alpha_1, \alpha_2, \ldots, \alpha_{n-1} \in \mathbb{R}$

Example: Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix} \tag{4}$$

Is $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ linearly independent?

Rank

The *column rank* of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the largest subset of columns of $\mathbf{A}$ that are linearly independent.

- The column rank is always $\leq n$.

The *row rank* of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the largest subset of rows of $\mathbf{A}$ that are linearly independent.

- The row rank is always $\leq m$.

Rank

The *column rank* of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the largest subset of columns of $\mathbf{A}$ that are linearly independent.

- The column rank is always $\leq n$.

The *row rank* of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the largest subset of rows of $\mathbf{A}$ that are linearly independent.

- The row rank is always $\leq m$.

n.b. Column rank is always equal to row rank. Thus, we refer to both as the *rank* of the matrix.

- For $\mathbf{A} \in \mathbb{R}^{m \times n}$, if $rank(\mathbf{A}) = \min(m, n)$, then $\mathbf{A}$ is said to be of *full rank*.
- For $\mathbf{A} \in \mathbb{R}^{m \times n}$, $rank(\mathbf{A}) = rank(\mathbf{A}^\top$.
- For $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$, $rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B}))$.
- For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$, $rank(\mathbf{A} + \mathbf{B}) \leq rank(\mathbf{A}) + rank(\mathbf{B})$

Matrix inverse

The *inverse* of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is denoted $\mathbf{A}^{-1}$, and is unique such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1} \tag{5}$$

Matrix inverse

The *inverse* of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is denoted $\mathbf{A}^{-1}$, and is unique such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1} \tag{5}$$

n.b. Not all matrices have inverses (e.g. $m \times n$ matrices).

Def:
A is *invertible* or *non-singular* if $\mathbf{A}^{-1}$ exists.
Otherwise, it is *non-invertible* or *singular*.

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
2. $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
3. $(\mathbf{A}^{-1})^{\top} = (\mathbf{A}^{\top})^{-1}$
    - This matrix is sometimes denoted $\mathbf{A}^{-\top}$

Orthogonal Matrices

Def:

- A vector $\mathbf{x} \in \mathbb{R}^n$ is *normalized* if $||\mathbf{x}||_2 = 1$
- Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are *orthogonal* if $\mathbf{x}^\top \mathbf{y} = 0$
- A square matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ is *orthogonal* or *orthonormal* if all its columns are:
    1. Orthogonal to each other
    2. Normalized

We therfore have that

$$\mathbf{U}^\top \mathbf{U} = \mathbf{I} = \mathbf{U}\mathbf{U}^\top \tag{6}$$

Orthogonal Matrices

Def:

- A vector $\mathbf{x} \in \mathbb{R}^n$ is *normalized* if $||\mathbf{x}||_2 = 1$
- Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are *orthogonal* if $\mathbf{x}^\top \mathbf{y} = 0$
- A square matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ is *orthogonal* or *orthonormal* if all its columns are:
    1. Orthogonal to each other
    2. Normalized

We therfore have that

$$\mathbf{U}^\top \mathbf{U} = \mathbf{I} = \mathbf{U}\mathbf{U}^\top \tag{6}$$

Another nice property:

$$||\mathbf{U}\mathbf{x}||_2 = ||\mathbf{x}||_2 \ \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{U} \in \mathbb{R}^{n \times n} \text{ orthogonal} \tag{7}$$

Range

Def:
The *span* of a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is

$$\text{span}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \alpha_i \in \mathbb{R} \right\} \tag{8}$$

Range

Def:
The *span* of a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is

$$\operatorname{span}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \alpha_i \in \mathbb{R} \right\} \tag{8}$$

n.b. If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is linearly independent, then $\operatorname{span}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}) = \mathbb{R}^n$.

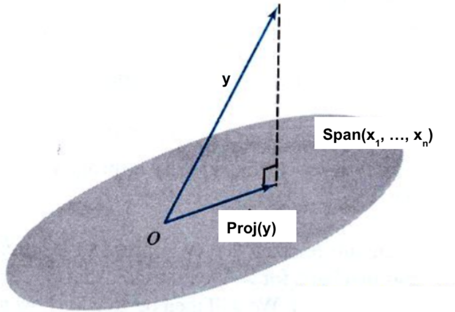
Example:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{9}$$

Projection

Def:
The *projection* of a vector $\mathbf{y} \in \mathbb{R}^m$ onto $\text{span}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}) = \mathbb{R}^n$ is

$$\text{Proj}(\mathbf{y}; \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}) = \underset{\mathbf{v} \in \text{span}(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\})}{\arg\min} ||\mathbf{y} - \mathbf{v}||_2 \qquad (10)$$



Range

Def:
The *range* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted $\mathcal{R}(\mathbf{A})$ is the span of the columns of $\mathbf{A}$, i.e.

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{v} \in \mathbb{R}^m : \mathbf{v} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\} \tag{11}$$

Assuming that $\mathbf{A}$ is full rank and $n < m$, the projection of $\mathbf{y} \in \mathbb{R}^m$ onto $\mathcal{R}(\mathbf{A})$ is

$$\begin{aligned} \mathrm{Proj}(\mathbf{y}; \mathbf{A}) &= \arg\min_{\mathbf{v} \in \mathcal{R}(\mathbf{A})} ||\mathbf{v} - \mathbf{y}||_2 && (12) \\ &= \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} && (13) \end{aligned}$$

Nullspace

Def:
The *nullspace* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted $\mathcal{N}(\mathbf{A})$ is the set of all vectors that equal 0 when multiplied by $\mathbf{A}$, i.e.

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = 0\} \tag{14}$$

Some properties:

- $\{w : w = u + v, u \in \mathcal{R}(\mathbf{A}^\top), v \in \mathcal{R}(\mathbf{A})\} = \mathbb{R}^n$
- $\mathcal{R}(\mathbf{A}^\top) \bigcap \mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$

This is referred to as *orthogonal complements*, denoted as $\mathcal{R}(\mathbf{A}^\top) = \mathcal{N}(\mathbf{A})^\perp$

Determinant

Def:
The *determinant* of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, denoted $|\mathbf{A}|$ or det $\mathbf{A}$ is a function det: $\mathbb{R}^{n \times n} \to \mathbb{R}$.

Let $\mathbf{A}_{\backslash i, \backslash j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the matrix that results from deleting the i^{th} row and j^{th} column. The general (recursive) formula for the determinant is

$$\begin{aligned} |\mathbf{A}| &= \textstyle\sum_{i=1}^{n}(-1)^{i+j} a_{ij} |\mathbf{A}_{\backslash i, \backslash j}| \quad (\forall j \in 1, ..., n) \\ &= \textstyle\sum_{j=1}^{n}(-1)^{i+j} a_{ij} |\mathbf{A}_{\backslash i, \backslash j}| \quad (\forall i \in 1, ..., n) \end{aligned} \tag{15}$$

Determinant

Given a matrix

$$\mathbf{A} = \begin{bmatrix} — & \mathbf{a}_1^\top & — \\ — & \mathbf{a}_2^\top & — \\ & \vdots & \\ — & \mathbf{a}_n^\top & — \end{bmatrix} \tag{16}$$

and a set $\mathbf{S} \subset \mathbb{R}^n$,

$$\mathbf{S} = \{\mathbf{v} \in \mathbb{R}^n : v = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i \text{ where } 0 \leq \alpha_i \leq 1, i = 1, ..., n\} \tag{17}$$

$|\mathbf{A}|$ is the volume of $\mathbf{S}$.

Determinant

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \tag{18}$$

Determinant

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \tag{18}$$

The matrix rows are:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \tag{19}$$

And $|\mathbf{A}| = -7$

Determinant

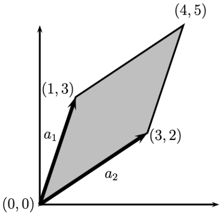
Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \tag{18}$$

The matrix rows are:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \tag{19}$$

And $|\mathbf{A}| = -7$



Determinant

Properties of determinants:

- For $\mathbf{A} \in \mathbb{R}^{n \times n}, |\mathbf{A}| = |\mathbf{A}^\top|$
- For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}, |\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
- For $\mathbf{A} \in \mathbb{R}^{n \times n}, |\mathbf{A}| = 0$ iff $\mathbf{A}$ is singular (i.e. non-invertible).
- For $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{A}$ non-singular, $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$

Quadratic form

Given $\mathbf{A} \in \mathbb{R}^{n \times n}$ and a vector $\mathbf{x} \in \mathbb{R}^n$, the *quadratic form* is the scalar value

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} = \sum_{i=1}^{n} x_i (\mathbf{A}\mathbf{x})_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} \mathbf{A}_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{A}_{ij} x_i x_j \quad (20)$$

Quadratic form

Some properties involving quadratic form:

- A symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ is *positive definite* if for a non-zero $\mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$
- A symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ is *positive semi-definite* if for a non-zero $\mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$
- A symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ is *negative definite* if for a non-zero $\mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top \mathbf{A}\mathbf{x} < 0$
- A symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ is *negative semi-definite* if for a non-zero $\mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top \mathbf{A}\mathbf{x} \leq 0$
- A symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ is *indefinite* if it is neither positive nor negative semidefinite

n.b. Positive definite and negative definite matrices always have full rank.

Eigenvalues & eigenvectors

Given $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an *eigenvalue* of $\mathbf{A}$ with corresponding *eigenvector* $\mathbf{x} \in \mathbb{C}^n$ if

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} : \mathbf{x} \neq 0 \tag{21}$$

n.b. The eigenvector is (usually) normalized to have length 1

Eigenvalues & eigenvectors

Given $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an *eigenvalue* of $\mathbf{A}$ with corresponding *eigenvector* $\mathbf{x} \in \mathbb{C}^n$ if

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} : \mathbf{x} \neq 0 \tag{21}$$

n.b. The eigenvector is (usually) normalized to have length 1

We can write all of the eigenvector equations simultaneously as

$$\mathbf{A}\mathbf{X} = \mathbf{X}\boldsymbol{\Lambda} \tag{22}$$

where

$$\mathbf{X} \in \mathbb{R}^{n \times n} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix}, \quad \boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_n) \tag{23}$$

This implies $\mathbf{A} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{-1}$

Eigenvalues & eigenvectors

Some properties:

- $tr\mathbf{A} = \sum_{i=1}^{n} \lambda_i$
- $|\mathbf{A}| = \prod_{i=1}^{n} \lambda_i$
- The rank of $\mathbf{A}$ is equal to the number of non-zero eigenvalues of $\mathbf{A}$.
- If $\mathbf{A}$ is non-singular, then $1/\lambda_i$ is an eigenvalue of $\mathbf{A}^{-1}$ with correspondng eigenvector $\mathbf{x}_i$, i.e. $\mathbf{A}^{-1}\mathbf{x}_i = (1/\lambda_i)\mathbf{x}_i$
- The eigenvalues of a diagonal matrix $D = diag(d_1, ..., d_n)$ are just its diagonal entries $d_1, ..., d_n$

Example: For $\mathbf{A} \in \mathbb{S}^n$ with ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$,

$$\max_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^\top \mathbf{A} \mathbf{x} \text{ subject to } ||\mathbf{x}||_2^2 = 1 \tag{24}$$

is solved with $\mathbf{x}_1$ corresponding to λ_1. Similarly, it is solved with $\mathbf{x}_n$ corresponding to λ_n.

Eigenvalues & eigenvectors

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ Find the eigenvalues & eigenvectors.

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ Find the eigenvalues & eigenvectors.

We want

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = 0 \tag{25}$$

Eigenvalues & eigenvectors

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ Find the eigenvalues & eigenvectors.

We want

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = 0 \tag{25}$$

We want $det(\mathbf{A} - \lambda \mathbb{I}) = 0$.

$$\begin{aligned} det(\mathbf{A} - \lambda \mathbb{I}) &= (1-\lambda)^2 - 2^2 = \lambda^2 - 2\lambda - 3 && (26) \\ &= (\lambda - 3)(\lambda + 1) && (27) \end{aligned}$$

$\therefore \ \lambda = 3, -1.$

Eigenvalues & eigenvectors

Finding the eigenvectors: calculating the null spaces of $(\mathbf{A} - \lambda\mathbf{I})$

$$\mathcal{N}(\mathbf{A} - 3\mathbf{I}) = \mathcal{N}\left(\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{28}$$

$$\mathcal{N}(\mathbf{A} + \mathbf{I}) = \mathcal{N}\left(\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{29}$$

Eigenvalues & eigenvectors

Finding the eigenvectors: calculating the null spaces of $(\mathbf{A} - \lambda \mathbf{I})$

$$\mathcal{N}(\mathbf{A} - 3\mathbf{I}) = \mathcal{N}\left(\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{28}$$

$$\mathcal{N}(\mathbf{A} + \mathbf{I}) = \mathcal{N}\left(\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{29}$$

Thus:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \mathbf{\Lambda} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \tag{30}$$

Singular Value Decomposition

SVD is a way of decomposing matrices.

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank r, $\exists$
$\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}, \mathbf{U} \in \mathbb{R}^{m \times m}, \mathbf{V} \in \mathbb{R}^{n \times m}$ $\ni$

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \tag{31}$$

Notes:

- $\boldsymbol{\Sigma}$ is a diagonal matrix with entries $\sigma_1, ..., \sigma_r > 0$ known as *singular values*.
- $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices.
- Common uses:
  - Least squares models
  - Range, rank, null space
  - Moore-Penrose inverse

Singular Value Decomposition

Some intuition:

$\mathbf{A} \in \mathbb{R}^{m \times n}$ can be thought of as a linear transformation, such that for $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} \tag{32}$$

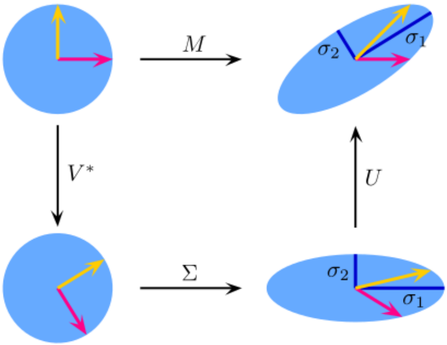
Singular Value Decomposition

Some intuition:

$\mathbf{A} \in \mathbb{R}^{m \times n}$ can be thought of as a linear transformation, such that for $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} \tag{32}$$

SVD can be thought of as breaking this into individual steps:



Matrix calculus

Given $f : \mathbb{R}^{m\times n} \to \mathbb{R}$, the *gradient* of f wrt $\mathbf{A} \in \mathbb{R}^{m\times n}$ is

$$\nabla_{\mathbf{A}} f(\mathbf{A}) \in \mathbb{R}^{m\times n} = \begin{bmatrix} \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{11}} & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{12}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{1n}} \\ \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{21}} & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{22}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{m1}} & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{m2}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}_{mn}} \end{bmatrix} \tag{33}$$

Some properties

- $\nabla_{\mathbf{x}}(f(\mathbf{x}) + g(\mathbf{x})) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \nabla_{\mathbf{x}} g(\mathbf{x})$
- For $c \in \mathbb{R}, \nabla_{\mathbf{x}}(c\, f(\mathbf{x})) = c\nabla_{\mathbf{x}}(f(\mathbf{x}))$

The Hessian

Given $f : \mathbb{R}^n \to \mathbb{R}$, the *Hessian* of f wrt $\mathbf{x} \in \mathbb{R}^n$ is

$$\nabla^2_{\mathbf{x}} f(\mathbf{x}) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix} \tag{34}$$

n.b. The Hessian is always symmetric, since $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$

Least squares

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m \ni b \notin \mathcal{R}(A)$, we want to find $\mathbf{x} \in \mathbb{R}^n$ as close as possible to $\mathbf{b}$ (via the Euclidean norm),

$$\begin{aligned} ||\mathbf{Ax} - \mathbf{b}||_2^2 &= (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) && (35) \\ &= \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - 2\mathbf{b}^\top \mathbf{Ax} + \mathbf{b}^\top \mathbf{b} && (36) \end{aligned}$$

Least squares

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m \ni b \notin \mathcal{R}(A)$, we want to find $\mathbf{x} \in \mathbb{R}^n$ as close as possible to $\mathbf{b}$ (via the Euclidean norm),

$$\begin{aligned} ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 &= (\mathbf{A}\mathbf{x} - \mathbf{b})^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) && (35) \\ &= \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{b}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{b} && (36) \end{aligned}$$

Taking the gradient wrt $\mathbf{x}$, we have

$$\begin{aligned} \nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{b}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{b}) &= \nabla_{\mathbf{x}}\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - \nabla_{\mathbf{x}} 2\mathbf{b}^\top \mathbf{A}\mathbf{x} + \nabla_{\mathbf{x}}\mathbf{b}^\top \mathbf{b} && (37) \\ &= \mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{A}^\top \mathbf{b} && (38) \end{aligned}$$

Least squares

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m \ni b \notin \mathcal{R}(A)$, we want to find $\mathbf{x} \in \mathbb{R}^n$ as close as possible to $\mathbf{b}$ (via the Euclidean norm),

$$
\begin{aligned}
||\mathbf{Ax} - \mathbf{b}||_2^2 &= (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) && (35) \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - 2\mathbf{b}^\top \mathbf{Ax} + \mathbf{b}^\top \mathbf{b} && (36)
\end{aligned}
$$

Taking the gradient wrt $\mathbf{x}$, we have

$$
\begin{aligned}
\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - 2\mathbf{b}^\top \mathbf{Ax} + \mathbf{b}^\top \mathbf{b}) &= \nabla_{\mathbf{x}} \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - \nabla_{\mathbf{x}} 2\mathbf{b}^\top \mathbf{Ax} + \nabla_{\mathbf{x}} \mathbf{b}^\top \mathbf{b} && (37) \\
&= \mathbf{A}^\top \mathbf{Ax} - 2\mathbf{A}^\top \mathbf{b} && (38)
\end{aligned}
$$

Setting this expression equal to zero and solving for $\mathbf{x}$ gives the normal equations,

$$
\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \qquad (39)
$$

References

Some textbooks on linear algebra:

- *Linear Algebra (Jim Hefferon)*
- *Introduction to Applied Linear Algebra (Boyd & Vandenberghe)*
- *Linear Algebra (Cherney, Denton et al.)*
- *Linear Algebra (Hoffman & Kunze)*
- *Fundamentals of Linear Algebra (Carrell)*
- *Linear Algebra (S. Friedberg A. Insel L. Spence)*