Basic Linear Algebra Refresher

Edith Zeng
CS 391L - Machine Learning
Dr. Dana Ballard
The University of Texas at Austin

March 6, 2019

Vector

Dot product is the sum of the product of two vectors' pairwise components:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$$

Angle between two vectors

$$cos\,\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{||\boldsymbol{x}||||\boldsymbol{y}||}$$

Length of a vector: $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}\boldsymbol{x}}$

Orthogonal vectors if their dot product is zero.
i.e. the angle $\theta = 90° \Leftrightarrow cos\,\theta = 0$.

Projection of vector $\boldsymbol{x}$ onto vector $\boldsymbol{y}$ results the vector

$$(\boldsymbol{x} \cdot \boldsymbol{y})\frac{\boldsymbol{y}}{||\boldsymbol{y}||}$$

Two graphs tagent at a given point $\Rightarrow$ normal vectors are parallel (scalar mutiples of each other).

Random Vectors

Density function

$$p(\boldsymbol{X}) = \lim_{\Delta \boldsymbol{x}_i \to 0} \frac{P\boldsymbol{X} \in \boldsymbol{I}}{\prod_i \Delta x_i} \text{ where } \boldsymbol{I} = \left\{\boldsymbol{X} : x_i < X_i \leq x_i + \Delta x_i\right\}$$

Face recognition Demean the average image from training set. Find eigenvectors. Project new observations to eigenspace and reconstruct (eg. KNN).

Mean vector $M = E[X] = \inf Xp(X)dX$
Covariance Matrix $\Sigma = E[(X - M)(X - M)^T]$
$= \frac{1}{N}\sum_{k=1}^{N}(\boldsymbol{X}^k - M)(\boldsymbol{X}^k - M)^T$

Linear Transformations

Scalar and vectors

$$T(k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + \ldots + k_r\boldsymbol{v}_r) = k_1T(\boldsymbol{v}_1) + k_2T(\boldsymbol{v}_2) + \ldots + k_rT(\boldsymbol{v}_r)$$

Matrix tmultiplication $\boldsymbol{A} = \boldsymbol{BC}$

$$a_{ij} = \sum_{k=1}^{N} b_{ik}c_{kj}$$

Matrix

Transpose

- $\boldsymbol{A}_{i,j}^T = \boldsymbol{A}_{j,i}$
- $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$
- $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
- $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
- If $\boldsymbol{A}$ has only real entries, $\boldsymbol{A}^T\boldsymbol{A}$ is positive semi-definite.

Symmetric: $\boldsymbol{A}^T = \boldsymbol{A}$
Orthogonal: $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$
Orthonormal: A transformation matrix A is orthonormal when $A^{-1} = A^T \Rightarrow AA^T = I$

Square Matrix Properties

Determinant

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow det(\boldsymbol{A}) = ad - bc$$

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$det(\boldsymbol{A}) = a \times det\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \times det\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \times det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

Properties

(1) If $\boldsymbol{B}$ is obtained by multiplying any row or column in $\boldsymbol{A}$ by a scalar λ, $det(\boldsymbol{B}) = \lambda det(\boldsymbol{A})$
(2) If $\boldsymbol{B}$ is obtained by interchanging two rows or columns in $\boldsymbol{A}$, $det(\boldsymbol{B}) = det(\boldsymbol{A})$
(3) $det(\boldsymbol{A}^T) = det(\boldsymbol{A})$
(4) $\boldsymbol{B}_{n\times n}$ and $\boldsymbol{A}_{n\times n} \Rightarrow det(\boldsymbol{AB}) = det(\boldsymbol{A})det(\boldsymbol{B})$

Inverse

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \boldsymbol{A}^{-1} = \frac{\begin{bmatrix} d & -b & \\ -c & & a \end{bmatrix}}{det(\boldsymbol{A})}$$

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

If $\boldsymbol{A}_{(m\times n)}$ and $\boldsymbol{B}_{(m\times n)}$ are both invertible:

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

If $\boldsymbol{A}$ is invertible:

- $\boldsymbol{A}^{-1} = \frac{adj(\boldsymbol{A})}{det(\boldsymbol{A})}$
- $det(\boldsymbol{A}^{-1}) = \frac{1}{det(\boldsymbol{A})}$

Invertability

$$det(\boldsymbol{A}) = \begin{cases} 0, & \text{singular} \Rightarrow \text{not invertible} \\ \text{else}, & \text{nonsingular} \Rightarrow \text{invertible} \end{cases}$$

SVD Matrix $\boldsymbol{A}$ is invertible iff all its singular values are non-zero.

$$A = U\Sigma V^T$$

$$\Sigma = U^T A V$$

$$\Sigma^{-1} = V^T A^{-1} U$$

Quadratic Forms and Positive Definitive Matrix

Quadratic Form of real, symmetric matrix $\boldsymbol{A}_{n\times n}$: $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$

Necessary and sufficient conditions for $\boldsymbol{A}$ to be positive definite

- All eigenvalues of $\boldsymbol{A}$ are positive, or
- determinant of every principal submatrix is positive.

Quadratic form $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$	Matrix
>0 for all $\boldsymbol{x} \neq 0$	positive definite
≥ 0	positive semi-definite
<0	negative definite
$\neq 0$	negative semi-definite
>0 for some $\boldsymbol{x}$ and < 0 for others	indefinite matrix

Properties of positive definite matrices

(1) If $\boldsymbol{A}_{n\times n}$ is symmetric and positive definite, a non-singular matrix $\boldsymbol{P}_{n\times n}$ can be found such that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{P}^T$
$\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \Lambda \Rightarrow A = (x\Lambda^{1/2}(X\Lambda^{1/2})^T)$ where $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of $\boldsymbol{A}$'s eigenvalues. $\boldsymbol{P} = \boldsymbol{X}\Lambda^{1/2}$.
(2) $\boldsymbol{A}$ is positive definite and $\rho(\boldsymbol{P}_{n\times n}) = m \Longrightarrow \boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ is positive definite.
(3) $\rho(\boldsymbol{A}) = n < m \Longrightarrow \boldsymbol{A}\boldsymbol{A}^T$ is positive semi-definite.
(4) $\rho(\boldsymbol{A}) = k < min(m, n) \Longrightarrow \boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^T$ are positive semi-definite.

Transformation of quadratic form to canonical form
Bivariate normal distribution and variance-covariance matrix

Properties of variance-covariance matrix

Definition The matrix whose i, j element is the covariance between the i-th and the j-th elements of the random vector. Potential values are specified by joint probability distribution. Symmetric.

$$\boldsymbol{Cov}[\boldsymbol{X},\boldsymbol{X}] = E[(\boldsymbol{X} - \mu_{\boldsymbol{x}})(\boldsymbol{X} - \mu_{\boldsymbol{x}})^T] = E[\boldsymbol{X}\boldsymbol{X}^T] - \mu_{\boldsymbol{X}}\mu_{\boldsymbol{X}}^T \text{ where } \mu_{\boldsymbol{x}} = E[\boldsymbol{X}]$$

Concavity/Convexity

For any $u \neq v$ in its domain and $0 < \theta < 1$,

$$\theta f(u) + (1-\theta)f(v) \begin{cases} \leq f[\theta u + (1-\theta)v] & \text{concave} \\ \geq f[\theta u + (1-\theta)v] & \text{convex} \end{cases}$$

i.e. draw the line segment between $f(u)$ and $f(v)$ and its relative position with f.
If f is differentiable,

$$\theta f(v) \begin{cases} \leq f(u) + \nabla f(u)(v-u) & \text{concave} \\ \geq f(u) + \nabla f(u)(v-u) & \text{convex} \end{cases}$$

i.e. in moving from one point to another on the surface of f, the height of the destination compared to the tangent line in that direction.

Quasiconcavity / quasiconvexity

Coordinate Systems

For the equation $\boldsymbol{Ax} = \mathbf{0}$ to have a nontrivial solution, the columns of $\boldsymbol{A}$ must have less than full rank.

For the equation $\boldsymbol{Ax} = \boldsymbol{c}$ to have a unique solution, $\boldsymbol{A}$ must have full rank: vector $\boldsymbol{c}$ must be expressed in terms of a linear combination of the columns in $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Matrix multiplications that change the magnitude of the vector but leaves the directions unchanged.

Definition If $\boldsymbol{A}$ is an $n \times n$ matrix, then a non-zero vector $\boldsymbol{x}$ in R^n is called an eigenvector of $\boldsymbol{A}$ if $\boldsymbol{Ax}$ is a scalar multiple of $\boldsymbol{x}$, that is

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$

for some scalar λ. The scalar is called an eigenvalue and $\boldsymbol{x}$ is an eigenvector corresponding to λ (column vector).

Characteristic equation: λ is an eigenvalue of $\boldsymbol{A}$ iff $det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$.

If $\boldsymbol{A}$ is a triangular matrix, the eigenvalues are on the main diagonal.

$\boldsymbol{x}_{n\times 1}$ is an eigenvector of $\boldsymbol{A}_{n\times n}$ iff $\boldsymbol{x}$ is a non-trivil solution of $(\lambda I - \boldsymbol{A})\boldsymbol{x} = \mathbf{0}$.

Finding eigenvalues and eigenvectors

$$\boldsymbol{A} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

eigenvalues:

$$det(\begin{bmatrix} \lambda - 3 & 0 \\ -8 & \lambda + 1 \end{bmatrix}) = 0 \ \Rightarrow (\lambda - 3)(\lambda + 1) = 0 \Rightarrow \lambda = 3 \text{ and } \lambda = -1$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Solve for $(\lambda I - \boldsymbol{A})\boldsymbol{x} = \mathbf{0}$, plug in $\lambda = 3$

$$\begin{bmatrix} 0 & 0 \\ -8 & 4 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

one solution:

$$\begin{bmatrix} \frac{1}{2}t \\ t \end{bmatrix}$$

Eigenvectors cannot be $\mathbf{0}$, but eigenvalues can be 0 (i.e. A is not invertible).

Eigenvectors corresponding to distinct eigenvalues of A are linearly independent.

If there are n linearly independent eigenvectors of $\boldsymbol{A}_{n\times n}$, the matrix is diagonalizable.

Spectral Theorem: if $A_{n\times n}$ is a symmetric real matrix, then it has n orthogonal eigenvectors. Proof: all eigenvalues—roots of the characteristic polynomial of A—are real numbers.

Theorem a matrix with all real eigenvalues and n orthogonal real eigenvectors iff it is a real symmetric matrix.

1. Q is the orthogonal matrix of eigenvectors.
2. $Q^T = Q^{-1} \Rightarrow A = Q\Lambda Q^T$ is real.
3. $A^T = (Q^T)^T\Lambda^T Q^T$ because $Q^{TT} = Q$.
4. The diagonal matrix is its own transpose, so A is symmetric.

Properties of Eigenvalues

- An eigenvalue matrix $\boldsymbol{A}$ is invariant under any orthogonal transformation.
- If all its eigenvalues are positive $\Rightarrow$ $\boldsymbol{A}$ is positive definite.
- $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^m) = \sum eigenvalues$ is invariant under any orthogonal transformation.
- $det(\boldsymbol{A}) = \prod eigenvalues$ is invariant under any orthogonal transformation.

PCA

Eigenvectors as basis Transforms the matrix into a diagonal matrix (eigenvalues). Reduce matrix multiplication in the old basis to scalar multiplication in the new basis.

Similar matrices and eigenvalue transformations: For a given matrix W, the transformation it defines can be simplified to a diagonal matrix of eigenvalues by transforming to coordinates that uses its eigenvectors as a basis.

1. For a given eigenvector $\boldsymbol{y}_i$, $W\boldsymbol{y}_i = \lambda\boldsymbol{y}_i$.
2. For the matrix Y with eigenvectors as columns, $WY = Y\Lambda$, where Λ is a diagonal matrix of the eigenvalues.
3. Pre-multiplying both sides by Y^{-1} gives $Y^{-1}WY = \Lambda$

ICA

vs. PCA: find new basis to represent the data. Different goal. Cocktail party problem with unmixing matrix $W = A^{-1}$.

Lagrange Multiplier

Goal: find a function's stationary points (derivatives = 0) several variables, subject to constraints. i.e. where the gradients points in the same direction as the constraints' gradients.

Advantages: solve optimization without explicit parameterization in terms of contraints.

Example - optimizing $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$:

(1) Solve the systems of equations

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$$
$$g(x, y, z) = k$$

which is equivalent to solving

$$\begin{cases} f(x) = \lambda g(x) \\ f(y) = \lambda g(y) \\ f(y) = \lambda g(y) \\ f(z) = \lambda g(z) \\ g(x, y, z) = k \end{cases}$$

(2) Plug in all solutions (x, y, z) into $f(x, y, z)$ (3) Find the minimum/maximum, provided that they exist and $\nabla g \neq 0$.

References

[1] Howard Anton and Chris Rorres. *Elementary Linear Algebra*. 10th ed. Wiley. 2011.

[2] Independent Component Analysis. `http://cs229.stanford.edu/notes/cs229-notes11.pdf`.

[3] `https://www.cs.utexas.edu/~dana/MLClass/`

[4] `https://www.cs.utexas.edu/~dana/MLClass/MT2013.pdf`

[5] `https://www.cs.utexas.edu/~dana/MLClass/MT2012.pdf`

[6] `https://www.cs.utexas.edu/~dana/MLClass/practice-midterm-1.pdf`

[7] `https://www.cs.utexas.edu/~dana/MLClass/practice-midterm-2.pdf`