CSI 436/536
Introduction to Machine Learning

Review of Linear Algebra and Calculus (2)

Professor Siwei Lyu
Computer Science
University at Albany, State University of New York
Matrices

- 2D tabular of numbers
- rank-2 tensor
- collection of column vectors, and their space
  \[
  X = \begin{pmatrix}
  x_1 & x_2 & \cdots & x_n \\
  \vdots & \vdots & \ddots & \vdots \\
  \end{pmatrix}, \quad \text{col}(X) = \text{span}(x_1, x_2, \cdots, x_n).
  \]
- collection of row vectors, and their space
- A linear transform
linear transforms and basis

• \( T(\mathbf{v}) = T(a_1\mathbf{e}_1 + \ldots + a_n\mathbf{e}_n) \)
  \[ = a_1T(\mathbf{e}_1) + \ldots + a_nT(\mathbf{e}_n) \]
  \[ = T\mathbf{a} \]

• \( T\mathbf{a} \) is matrix-vector product
  • \( T \) is a matrix each column corresponding to \( T(\mathbf{e}_i) \)
  • \( \mathbf{a} \) is a vector containing all the values of \( a_i \)

• so any linear transform is equivalent to a matrix and vice versa
linear transforms

- definition: T is a mapping between vector spaces, and satisfy: \(aT(x) + bT(y) = T(ax + by)\)
- two equivalent effects
  - move all the points (active transform)
  - move the basis (inactive transform)
  - translation \(T(x) = x+c\) is not linear (but can be made so)
- basic “linear transforms
  - rotation
  - scaling
    - isometric scaling
    - anisotropic scaling
  - rotation + scaling = shear transform
  - rotation + isometric scaling = conformal transform
  - rotation + translation = rigid transform
  - rotation + scaling + translation = affine transform
eigenvalue and eigenvector

- An eigenvector of a square matrix $T$ (equivalently a linear transform) is a non-zero complex vector $v$ which $T$ sends to a complex multiple (the eigenvalue) of itself: $Tv = \lambda v$

- for an $n \times n$ matrix there are exactly $n$ eigenvalues (counting zero and complex numbers)

- determinant is a polynomial of $n$-degree (Caley-Hamilton theorem)
how to solve eigenvalue problem

• solve: \(Tv = \lambda v\)
  • equivalently, we write \((T - \lambda I)v = 0\)
  • so that if \((\lambda, v)\) are eigenvalue-eigenvector of \(T\), matrix \((T - \lambda I)\) is singular
  • or we solve \(\det(T - \lambda I) = 0\)
    • this is known as the characteristic polynomial of matrix \(T\)

\[
|A - \lambda \cdot I| = \begin{vmatrix} 0 & 1 \\ -2 & -3 \end{vmatrix} - \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} = 0
\]

\[
\begin{vmatrix} -\lambda & 1 \\ -2 & -3 - \lambda \end{vmatrix} = \lambda^2 + 3\lambda + 2 = 0
\]
an application of eigenvalues

Fibonacci number is defined as follows:

\[ F_1 = 1, \ F_2 = 1, \ F_{n+1} = F_n + F_{n-1}, \ n = 3, \cdots \]

\[ f_{k+2} = f_{k+1} + f_k \]

\[
\begin{bmatrix}
  f_{k+1} \\
  f_k 
\end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} v_k
\]

\[
v_{k+1} = \begin{bmatrix} f_{k+2} \\ f_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} v_k
\]

\[ A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \]

\[
|A - \lambda I| = \begin{vmatrix} 1 - \lambda & 1 \\ 1 & 1 - \lambda \end{vmatrix}
\]

\[
\lambda^2 - \lambda - 1 = 0
\]

\[ \lambda_1 = \frac{1 + \sqrt{5}}{2} \]

\[ \lambda_2 = \frac{1 - \sqrt{5}}{2} \]

\[ v_{100} = A^{100} v_0 \]

\[ A^{100} = S \Lambda^{100} S^{-1} \]
Matrix trace & determinant

- trace = sum of eigenvalues
- Trace is a linear function of matrix
- property: $\text{tr}(AB) = \text{tr}(BA^T)$
- determinant = product of eigenvalues

- $\det(aA) = a^n \det(A)$, $\det(AB) = \det(A) \det(B)$, $\det(A^{-1}) = \det(A)^{-1}$, $\det(e^A) = e^{\text{tr}(A)}$
- If $A$ is not invertible, then $\det(A) = 0$, and vice versa
spectral theorem

- A real symmetric matrix can be decomposed as

\[ A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \lambda_1 x_1 x_1^T + \cdots + \lambda_n x_n x_n^T \]

- \( x_1, x_2, \ldots x_n \) are eigenvectors that can be chosen as real vectors
- \( \lambda_1, \lambda_2, \ldots \lambda_n \) are real eigenvalues
- Real symmetric matrix is diagonalizable with orthonormal matrices as \( A = U \Lambda U^T \)
Some notes

• not every square matrix can be diagonalized, but every square matrix has a Jordan standard form
  • Example: rank-1 matrix $uv^T$ when $u^Tv = 0$

• All normal matrices $A^*A = AA^*$ can be diagonalized
  • symmetric matrices
  • Hermitian matrices
  • Orthogonal matrices

• Every real rectangular matrix can be diagonalized using the singular value decomposition (SVD) as $A = U\Lambda V^T$, where $U$ and $V$ are orthonormal matrices, $\Lambda$ is a rectangular diagonal matrix with positive diagonals (the singular values)
positive (semi) definite matrices

• $v^TAv$ is the quadratic form of vector $v$
  • $A$ is p.d. if for any non-zero $v$, $v^TAv > 0$
    • $A$ has all positive eigenvalues
  • $A$ is p.s.d. if for any non-zero $v$, $v^TAv \geq 0$
    • $A$ has all nonnegative eigenvalues
equivalence of PD

- A real symmetric matrix is positive definite (p.d.) if and only if all eigenvalues are positive.
  - $\implies$: pick any eigenvalue, eigenvector
  - $\impliedby$: use spectral theorem
two important p.s.d. matrices

- for data matrix $X$ (column vectors as data)
  - Gram (inner product) matrix: $G = X^TX$
  - correlation (covariance, outer product) matrix: $C = XX^T$
- $G$ and $C$ are both positive definite matrices
- $G$ and $C$ share the same non-zero eigenvalues
  - if $\lambda$ and $v$ are eigenvalue and the corresponding eigenvector of $X^TX$, we have $X^TXv = \lambda v$
  - then we have $XX^TXv = \lambda Xv$, or $XX^Tu = \lambda Xu$, where $u = Xv$
- $G$ and $C$’s eigenvectors are related by $X$
- They are dual to each other