Just Enough Spectral Theory

Ng Yen Kaow

Notations (Important)

- A vector is by default a column
  - For vectors x and y, their inner (or dot) product, $\langle x, y \rangle = x^{\mathrm{T}} y$
    - $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle = x^T y + z^T y$
  - Beware: some texts use row vectors and $\langle x, y \rangle = x y^{\mathrm{T}}$

- For a matrix an example is a row
  - An example (or datapoint) is a row x_i while each feature is a columns
    - Features are like fixed columns in a spreadsheet
  - For matrices X and Y, $\langle X, Y \rangle = X Y^{\mathrm{T}}$ or $\sum_i (x_i y_i^{\mathrm{T}})$
  - Beware: some texts use column for examples and let $\langle X, Y \rangle = X^{\mathrm{T}} Y$

- So it's $x^{\mathrm{T}} x$, $x^{\mathrm{T}} M x$, but $X X^{\mathrm{T}}$ and $Q \Lambda Q^{\mathrm{T}}$

Outer product

- The outer product of two vectors x and y is a matrix M where the $M_{ij} = x_i y_j$
  e.g. $\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix}$
- The outer product (or Kronecker product) of two matrices is a tensor
  - We don't deal with tensors yet
- Common uses of outer products
  - Denote pairwise inner product matrix, $xx^{\mathrm{T}} = \begin{pmatrix} x_1x_1 & x_1x_2 & \dots \\ x_2x_1 & x_2x_2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$
  - Denote matrix of all ones, $\mathbf{1}\mathbf{1}^{\mathrm{T}} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$

More notations

- Conventions
  - x_i from a matrix is by default a row vector
  - x_i from a vector is a scalar
  - x_{ij} from a matrix is a scalar
  - x, u_i (all other vectors) are by default column vectors
- Common expansions

$$xy^{\mathrm{T}} = \sum_i x_i y_i \qquad (XY)_{ij} = \sum_k x_{ik} y_{kj}$$

$$\left(x^{\mathrm{T}} y\right)_{ij} = x_i y_j \qquad \left(XY^{\mathrm{T}}\right)_{ij} = x_i y_j^{\mathrm{T}} = \sum_k x_{ik} y_{jk}$$

$$x^{\mathrm{T}} M y = \sum_{ij} m_{ij} x_i y_j \qquad \left(X^{\mathrm{T}} Y\right)_{ij} = \sum_k x_{ki} y_{kj}$$

$X^{\mathrm{T}} X = \sum_i x_i^{\mathrm{T}} x_i$ (used in kernel PCA)

Python call for inner product

- Inner products are performed with `np.dot()`
  - When called on two arrays, the arrays are **automatically** oriented to perform inner product
    - Note that `[[1],[1]]` is a 1×2 matrix
  - When called on an array `x` and a matrix `X`, the array is **automatically** read as a row for `np.dot(x,X)`, and column for `np.dot(X,x)` to perform inner product
  - When called on two matrices, make sure that the matrices are oriented correctly, or you will get $X^{\mathrm{T}}X$ when you want XX^{T}
  - Impossible to get outer product with `np.dot()`
- If you write `x*y` or `X*Y`, what you get is an element-wise multiplication

Eigenvectors and eigenvalues

- Only concerned with square matrices
  - Most matrices we consider are furthermore **symmetric** and of only **real** values
- A eigenvector for a square matrix M is vector u where $Mu = \lambda u$
  - u is invariant under transformation M
  - The scaling factor λ is a eigenvalue
  - Use u to denote a column vector even when multiple u_i are collected into a matrix $U = [u_1 \quad \ldots \quad u_k]$

$Mu = \lambda u$ is a system of equations

- An equation such as $Mu = \lambda u$ actually states n linear equations, namely $\forall i, \sum_j m_i u_j = \lambda u_i$
  - For example
    $$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \lambda \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$
    states the two equations
    $$m_{11}u_1 + m_{12}u_2 = \lambda u_1$$
    $$m_{21}u_1 + m_{22}u_2 = \lambda u_2$$
- This is important when manipulating equation by multiplying with other matrix/vector
  - For example when $Mu = \lambda u$ is multiplied from the left by u^{T}, the resultant $u^{\mathrm{T}}Mu = \lambda u^{\mathrm{T}}u$ becomes only one equation, that is, $\sum_{ij} u_i m_{ij} u_j = \lambda \sum_{ij} u_i u_j$

Eigendecomposition

- A eigendecomposition of matrix M is
$$M = Q\Lambda Q^{-1}$$
where Λ is diagonal, and Q contains (not necessarily orthogonal) eigenvectors of M
  - Any normal M can be eigendecomposed
  - **The set of eigenvalues for M is unique**
  - **There can be different eigenvectors of the same eigenvalue** (hence not unique)
    - **For real symmetric M, eigenvectors that correspond to distinct eigenvalues are (chosen to be) orthogonal**

Orthogonal eigendecomposition

- For real symmetric M, can choose Q to be orthogonal matrix (proof omitted)
- For square matrix Q, the following are equivalent (proof next slide)
  1. Q is an orthogonal matrix
  2. $Q^{\mathrm{T}}Q = I$
  3. $QQ^{\mathrm{T}} = I$
  - Corollary. $Q^{\mathrm{T}}Q = I \Rightarrow Q^{\mathrm{T}}QQ^{-1} = Q^{-1}$

    $\Rightarrow Q^{\mathrm{T}} = Q^{-1}$
- By default the **eigendecomposition of real symmetric matrix** M **is** $M = Q\Lambda Q^{\mathrm{T}}$

Orthogonal matrix property

- For square matrix Q, the following are equivalent
  1. Q is orthogonal matrix
  2. $Q^{\mathrm{T}}Q = I$
  3. $QQ^{\mathrm{T}} = I$

  2⇔1 Let u_i be the column vectors of A

$$Q^{\mathrm{T}}Q = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} [u_1 \quad \dots \quad u_n] = \begin{bmatrix} u_1u_1 & \dots & u_1u_n \\ \vdots & \ddots & \vdots \\ u_nu_1 & \dots & u_nu_n \end{bmatrix}$$

$$\begin{bmatrix} u_1u_1 & \dots & u_1u_n \\ \vdots & \ddots & \vdots \\ u_nu_1 & \dots & u_nu_n \end{bmatrix} = I \text{ implies } u_iu_j = 0 \text{ for } i \neq j$$

Eigenspace

- The eigenspace of a matrix M is the set of all the vectors u that fulfills $Mu = \lambda u$
  - The rank of M is its number of non-zero λ
- A eigenbasis of a $n \times n$ matrix M is a set of n orthogonal eigenvectors of M (including those with zero eigenvalues)
  - Any datapoint x_i in M can be written as a linear combination of the eigenbasis, $x_i = \sum_j \langle x_i, u_j \rangle u_j$
  - Any eigenvector u_i for M can be written as a linear combination of the datapoints x_i, by solving the system of equations $x_i = \sum_j \langle x_i, u_j \rangle u_j$

Rayleigh Quotient

- $\frac{u^{\mathrm{T}}Mu}{u^{\mathrm{T}}u}$ is called the **Rayleigh quotient**
- Let $\lambda_1, \ldots, \lambda_n$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of M
- **Min-max Theorem** (simplified)
  - Maximum of the Rayleigh quotient,
$$\max_{\|u\|=1} \frac{u^{\mathrm{T}}Mu}{u^{\mathrm{T}}u} = \lambda_1$$
  - Minimum of the Rayleigh quotient,
$$\min_{\|u\|=1} \frac{u^{\mathrm{T}}Mu}{u^{\mathrm{T}}u} = \lambda_n$$

Proof of min-max theorem

- **Find stationary points of $\frac{u^{\mathrm{T}}Mu}{u^{\mathrm{T}}u}$**
- Letting $u' = cu$ does not change $\frac{u^{\mathrm{T}}Mu}{u^{\mathrm{T}}u}\left(= \frac{u'^{\mathrm{T}}Mu'}{u'^{\mathrm{T}}u'}\right)$
  - Hence consider only unit u
  - Maximize $u^{\mathrm{T}}Mu$ subject to $u^{\mathrm{T}}u = 1$
- Use Lagrangian to add $u^{\mathrm{T}}u = 1$ constraint

$$\mathcal{L}(u,\lambda) = u^{\mathrm{T}}Mu + \lambda(u^{\mathrm{T}}u - 1)$$

$$\frac{\partial\mathcal{L}}{\partial u} = u^{\mathrm{T}}(M + M^{\mathrm{T}}) + 2\lambda u^{\mathrm{T}} = 0$$

$$\frac{\partial\mathcal{L}}{\partial \lambda} = u^{\mathrm{T}}u - 1 = 0$$

$$u^{\mathrm{T}}(M + M^{\mathrm{T}}) = -2\lambda u^{\mathrm{T}} \Rightarrow (M + M^{\mathrm{T}})u = -2\lambda u$$

Since M is symmetric, $2Mu = -2\lambda u$

$\Rightarrow Mu = \tilde{\lambda}u$ where $\tilde{\lambda} = -2\lambda$

Matrix differentiation*

$$\frac{\partial x^{\mathrm{T}}Mx}{\partial x} = x^{\mathrm{T}}(M + M^{\mathrm{T}})$$

$$\frac{\partial x^{\mathrm{T}}x}{\partial x} = 2x^{\mathrm{T}}$$

- **Stationary points are solutions of $Mu = \tilde{\lambda}u$**

* https://www.comp.nus.edu.sg/~CS5240/lecture/matrix-diff.pdf

Eigendecomposition applications

- Matrix inverse
- Matrix approximation
- Matrix factorization
  - Multidimensional Scaling
- Minimization or maximization through the Rayleigh Quotient
  - PCA
    - Max of covariance matrix
  - Spectral clustering
    - Min of graph Laplacian

Singular Value Decomposition

- Any matrix can be singular value decomposed
- $M = U\Sigma \mathrm{V}^*$
  - M is $m \times n$ matrix
  - U is an $m \times m$ unitary (orthogonal) matrix
  - Σ is an $m \times n$ diagonal matrix
  - V is an $n \times n$ unitary matrix

- For a real M, $V^* = V^{\mathrm{T}}$ (and $U = U^{\mathrm{T}}$) hence

$$M = U\Sigma V^{\mathrm{T}}$$

SVD applications

- Solving linear equations
- Linear regression
- Pseudoinverse
- Kabsch algorithm
- Matrix approximation
- As a eigendecomposition (see next slide)

SVD and eigendecomposition

- SVD is a eigendecomposition but not of M
  - Given an SVD of $M = U\Sigma V^{\mathrm{T}}$
  - Then, clearly
    - $M^{\mathrm{T}}M = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = V\left(\Sigma^{\mathrm{T}}\Sigma\right)V^{\mathrm{T}}$
    - $MM^{\mathrm{T}} = U\Sigma V^{\mathrm{T}}V\Sigma^{\mathrm{T}}U^{\mathrm{T}} = U\left(\Sigma^{\mathrm{T}}\Sigma\right)U^{\mathrm{T}}$
  - Hence V is the eigenbasis of $M^{\mathrm{T}}M$ and U is the eigenbasis of MM^{T} respectively
  - That is, U and V are eigenbases of the squared matrices of M
    - However the eigenbasis of $M^{\mathrm{T}}M$ and MM^{T} are in general not the eigenbasis of M

Special Matrices

- Three types of matrices lead to most of the results
  - Covariance ($A^T A$ for column centered A)
    $\Rightarrow$ Principal Component Analysis
  - Gramian (AA^T for column centered A)
    $\Rightarrow$ Multidimensional Scaling
    $\Rightarrow$ Kernel Method
  - Graph Laplacian (AA^T for incidence matrix A)
    $\Rightarrow$ Spectral Clustering