COMP3670 2021 Theory Assignment 4

Yuxuan Lin – u6828533@anu.edu.au

October 24, 2021

Introduction to Machine Learning

By turning in this assignment, I agree by the ANU honor code and declare that all of this is my own work.

Exercise 1

1. ***Proof by contradiction:*** Assume that there exists an eigenvalue λ of $\boldsymbol{A}$, such that $\lambda = 0$, and there exists an $\boldsymbol{x} \neq \boldsymbol{0}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} = 0 \cdot \boldsymbol{x} = \boldsymbol{0}$. Since $\boldsymbol{A}$ is invertible, we can multiply $\boldsymbol{A}^{-1}$ to both sides of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, then we derive $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{0}$, thus $\boldsymbol{x} = \boldsymbol{0}$. However, this contradicts our earlier assumption that $\boldsymbol{x} \neq \boldsymbol{0}$, so our assumption must have been wrong. Therefore, all the eigenvalues of $\boldsymbol{A}$ are non-zero.

2. For any eigenvalue λ of $\boldsymbol{A}$, we have $\lambda \neq 0$ from the conclusion in 1. Then, there exists an $\boldsymbol{x} \neq \boldsymbol{0}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. Since $\boldsymbol{A}$ is invertible, we can multiply $\boldsymbol{A}^{-1}$ to both sides of $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, then we derive $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\lambda\boldsymbol{x}$. For this equation, $LHS = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}$, $RHS = \lambda\boldsymbol{A}^{-1}\boldsymbol{x}$, then we derive $\lambda\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{x}$. Since $\lambda \neq 0$, we can multiply λ^{-1} to both sides of $\lambda\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{x}$, then we derive $\boldsymbol{A}^{-1}\boldsymbol{x} = \lambda^{-1}\boldsymbol{x}$ for $\boldsymbol{x} \neq \boldsymbol{0}$. By eigenvalue definition, we have proved that λ^{-1} is an eigenvalue of $\boldsymbol{A}^{-1}$.

Exercise 2

Proof: We will prove by induction that, for all $n \in \mathbb{Z}_+$, $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{B}^n$ with eigenvalue λ^n, that is

$$\boldsymbol{B}^n\boldsymbol{x} = \lambda^n\boldsymbol{x} \tag{1}$$

Base case: When $n = 1$, we have $\boldsymbol{B}\boldsymbol{x} = \lambda\boldsymbol{x}$ by definitions of eigenvalue and eigenvector, thus assumption is true for $n = 1$.

Induction step: Let $k \in \mathbb{Z}_+$ be given and suppose assumption is true for $n = k$. Then for $n = k+1$, we derive

$$\begin{aligned}
\boldsymbol{B}^{k+1}\boldsymbol{x} &= \boldsymbol{B}(\boldsymbol{B}^k\boldsymbol{x}) \ldots \text{associativity} \\
&= \boldsymbol{B}(\lambda^k\boldsymbol{x}) \ldots . \text{induction assumption} \\
&= \lambda^k(\boldsymbol{B}\boldsymbol{x}) \ldots . \text{associativity} \\
&= \lambda^k(\lambda\boldsymbol{x}) . \ldots . \text{base case} \\
&= \lambda^{k+1}\boldsymbol{x}
\end{aligned}$$

Thus, (1) holds for $n = k+1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, we have proved that for all $n \in \mathbb{Z}_+$, $\boldsymbol{B}^n\boldsymbol{x} = \lambda^n\boldsymbol{x}$, thus $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{B}^n$ with eigenvalue λ^n.

Exercise 3

1. ***Proof by contradiction:*** Assume that $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ is linearly dependent, then there exists some p such that $1 \leq p < n$, $\boldsymbol{x}_{p+1} \in \text{span}\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\}$ and $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\}$ is linearly independent.
   Since $\boldsymbol{x}_{p+1} \in \text{span}\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\}$, then $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p, \boldsymbol{x}_{p+1}\}$ is linearly dependent by definition. Hence, there exist $\sigma_1, ..., \sigma_{p+1}$, with at least one $\sigma_n \neq 0$, such that

   $$\sigma_1\boldsymbol{x}_1 + ... + \sigma_p\boldsymbol{x}_p + \sigma_{p+1}\boldsymbol{x}_{p+1} = \boldsymbol{0} \tag{2}$$

   Multiply $\boldsymbol{A}$ to both sides of the equation, we have

   $$\begin{aligned}
   \boldsymbol{A}(\sigma_1\boldsymbol{x}_1 + ... + \sigma_p\boldsymbol{x}_p + \sigma_{p+1}\boldsymbol{x}_{p+1}) &= \sigma_1\boldsymbol{A}\boldsymbol{x}_1 + ... + \sigma_p\boldsymbol{A}\boldsymbol{x}_p + \sigma_{p+1}\boldsymbol{A}\boldsymbol{x}_{p+1} \ldots \text{distributivity} \\
   &= \sigma_1\lambda\boldsymbol{x}_1 + ... + \sigma_p\lambda\boldsymbol{x}_p + \sigma_{p+1}\lambda\boldsymbol{x}_{p+1} \ldots \text{eigenvalue definition} \\
   &= \boldsymbol{0}
   \end{aligned}$$

   We derive

   $$\sigma_1\lambda\boldsymbol{x}_1 + ... + \sigma_p\lambda\boldsymbol{x}_p + \sigma_{p+1}\lambda\boldsymbol{x}_{p+1} = \boldsymbol{0} \tag{3}$$

   If we perform equation transformation: $(3) - \lambda_{p+1} \cdot (2)$, we derive

   $$\sigma_1(\lambda_1 - \lambda_{p+1})\boldsymbol{x}_1 + ... + \sigma_p(\lambda_p - \lambda_{p+1})\boldsymbol{x}_p + \sigma_{p+1}(\lambda_{p+1} - \lambda_{p+1})\boldsymbol{x}_{p+1} = \boldsymbol{0}$$

   Then

   $$\sigma_1(\lambda_1 - \lambda_{p+1})\boldsymbol{x}_1 + ... + \sigma_p(\lambda_p - \lambda_{p+1})\boldsymbol{x}_p = \boldsymbol{0}$$

   Since $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\}$ is linearly independent, then the sum of coefficients should be $\boldsymbol{0}$. Notice that eigenvalues $\{\lambda_1, ..., \lambda_n\}$ are distinct, we can conclude that $\sigma_1 = ... = \sigma_p = 0$.

Take this conclusion into equation (2) we derive that: there exist $\sigma_1, ..., \sigma_{p+1}$, with at least one $\sigma_n \neq 0$, such that

$$\sigma_{p+1}\boldsymbol{x}_{p+1} = \mathbf{0} \tag{4}$$

Since $\boldsymbol{x}_{p+1}$ is an eigenvector, $\boldsymbol{x}_{p+1} \neq \mathbf{0}$, then $\sigma_{p+1} = 0$, $\sigma_1 = ... = \sigma_{p+1} = 0$, which contradicts our earlier assumption that at least one $\sigma_n \neq 0$, so our assumption must have been wrong. Therefore, we reject the assumption, and we have proved that $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ is linearly independent.

2. Eigenvalues of an $n \times n$ matrix $\boldsymbol{B}$ can be seen as the roots of the nth degree characteristic polynomial $p_{\boldsymbol{B}(\lambda)} = \det(\boldsymbol{B} - \lambda\boldsymbol{I})$ of $\boldsymbol{B}$. From the fundamental algebra theorem we know that this nth degree polynomial has exactly n roots, which indicates there are n corresponding eigenvalues for $\boldsymbol{B}$, either distinct or duplicated. Hence, there can be at most n distinct eigenvalues for $\boldsymbol{B}$.

Exercise 4

1. ***Proof:***
   We will prove by induction that, for all $A \in \mathbb{R}^{n\times n}$, $n \in \mathbb{Z}_+$,

$$\det(A) = \det\left(A^T\right) \tag{5}$$

Expansion along row j, we derive from (5) that

$$\sum_{n=1}^{k}(-1)^{n+j}a_{jn}\det(A_{j,n}) = \sum_{n=1}^{k}(-1)^{n+j}a_{jn}\det\left(A_{j,n}^T\right) \tag{6}$$

Base case:
When $n = 1$, $A \in \mathbb{R}^{1\times 1}$,$A = A^T$, $\det(A) = \det\left(A^T\right)$, assumption is true for $n = 1$.

Induction step:
Let $k \in \mathbb{Z}_+$ be given and suppose assumption is true for $n = k$.

Then for $n = k+1$, $A \in \mathbb{R}^{(k+1)\times(k+1)}$, we derive

$$\begin{aligned}
\det(A) &= \sum_{n=1}^{k+1}(-1)^{n+j}a_{jn}\det(A_{j,n}) \\
&= \sum_{n=1}^{k}(-1)^{n+j}a_{jn}\det(A_{j,n}) + (-1)^{(k+1)+j}a_{j(k+1)}\det(A_{j,k+1}) \\
&= \sum_{n=1}^{k}(-1)^{n+j}a_{jn}\det\left(A_{j,n}^T\right) + (-1)^{(k+1)+j}a_{j(k+1)}\det(A_{j,k+1})\ldots\text{equation (6)} \\
&= \sum_{n=1}^{k}(-1)^{n+j}a_{jn}\det\left(A_{j,n}^T\right) + (-1)^{(k+1)+j}a_{j(k+1)}\det\left(A_{j,k+1}^T\right)\ldots\text{base case} \\
&= \sum_{n=1}^{k+1}(-1)^{n+j}a_{jn}\det\left(A_{j,n}^T\right) \\
&= \det\left(A^T\right)
\end{aligned}$$

Thus, (5) holds for $n = k+1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, we have proved that $\det\left(A^T\right) = \det(A)$.

2. From the equation **(4.8)** of *mml-book* p.101, for an upper triangular matrix A, its determinant $\det(A)$ is equal to the product of all the diagonal elements of A. From the definition of identity matrix, for an $n \times n$ identity matrix I_n, there are n of 1 on the diagonal, known as $a_{ii} = 1$ for $i = 1, ..., n$, then we derive

$$\det(I_n) = \prod_{i=1}^{n} a_{ii} = 1^n = 1$$

Hence, we have proved that $\det(I_n) = 1$ where I_n is the $n \times n$ identity matrix.

Exercise 5

By definition of eigenvalue/eigenvector, we have

$$Av_1 = \lambda_1 v_1 , \quad Av_2 = \lambda_2 v_2$$

We'll first prove $\lambda_1 v_1^T v_2 = \lambda_2 v_1^T v_2$:

$$
\begin{aligned}
LHS &= \lambda_1 v_2^T v_1 && \text{... identity} \\
&= v_2^T (\lambda_1 v_1) && \text{... associativity} \\
&= v_2^T (A v_1) \\
&= (A v_1)^T v_2 && \text{... identity} \\
&= v_1^T A^T v_2 && \text{... transpose property} \\
&= v_1^T (A v_2) && \text{... } A \text{ is symmetric} \Rightarrow A^T = A \\
&= v_1^T (\lambda_2 v_2) \\
&= \lambda_2 v_1^T v_2 && \text{... associativity} \\
&= RHS
\end{aligned}
$$

Hence, $\lambda_1 v_1^T v_2 = \lambda_2 v_1^T v_2$

$\Rightarrow (\lambda_1 - \lambda_2) v_1^T v_2 = 0$

Notice that $\lambda_1 \neq \lambda_2$, then $v_1^T v_2 = 0$.

Recall that the inner product $\langle v_1, v_2 \rangle = v_1^T v_2$,

then $\langle v_1, v_2 \rangle = 0$,

which indicates v_1, v_2 are orthogonal.

Exercise 6.

1. Eigenvalues.

$$P_A(\lambda) = \det(A-\lambda I)$$

$$= \det\left(\begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)$$

$$= \begin{vmatrix} -1-\lambda & 2 \\ 3 & 4-\lambda \end{vmatrix}$$

$$= (\lambda-5)(\lambda+2)$$

Let $P_A(\lambda) = 0 \Rightarrow (\lambda-5)(\lambda+2) = 0 \Rightarrow \lambda_1 = 5,\ \lambda_2 = -2$

Hence, the eigenvalues are $\lambda_1 = 5$, $\lambda_2 = -2$.

2. Eigenvectors and Eigenspaces

For $\lambda_1 = 5$, ~~$A-\lambda_1 I$~~ $A - \lambda_1 I = \begin{bmatrix} -6 & 2 \\ 3 & -1 \end{bmatrix}$

$$\begin{bmatrix} -6 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} -6x_1 + 2x_2 = 0 \\ 3x_1 - x_2 = 0 \end{cases} \Rightarrow x_2 = 3x_1$$

Hence, $E_5 = \text{span}\left[\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right]$.

For $\lambda_2 = -2$, $A - \lambda_2 I = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2x_2 = 0 \\ 3x_1 + 6x_2 = 0 \end{cases} \Rightarrow x_1 = -2x_2$$

Hence, $E_{-2} = \text{span}\left[\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right]$.

In sum, $E_5 = \text{span}\left[\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right]$ and $E_{-2} = \text{span}\left[\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right]$.

3. The set of all eigenvectors of A is :

$$\text{span}\left[\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}-2\\1\end{bmatrix}\right].$$

For $\begin{bmatrix}1\\3\end{bmatrix}$ and $\begin{bmatrix}-2\\1\end{bmatrix}$, if there exists λ_1 and λ_2 that

$$\lambda_1\begin{bmatrix}1\\3\end{bmatrix} + \lambda_2\begin{bmatrix}-2\\1\end{bmatrix} = 0$$

$$\Rightarrow \begin{cases}\lambda_1 - 2\lambda_2 = 0\\ 3\lambda_1 + \lambda_2 = 0\end{cases} \Rightarrow \lambda_1 = \lambda_2 = 0$$

Hence, $\begin{bmatrix}1\\3\end{bmatrix}$ and $\begin{bmatrix}-2\\1\end{bmatrix}$ are linearly independent.

Then, the set of all eigenvectors of A spans $\mathbb{R}^2$.

4. From mml-book P.116,

the columns of P contain the eigenvectors of A,

D is diagonal matrix whose diagonal entries are eigenvalues of A.

Hence, we construct : $P = \begin{bmatrix}1 & -2\\3 & 1\end{bmatrix}$, $D = \begin{bmatrix}5 & 0\\0 & -2\end{bmatrix}$.

Validation : $PDP^{-1} = \begin{bmatrix}1 & -2\\3 & 1\end{bmatrix}\cdot\begin{bmatrix}5 & 0\\0 & -2\end{bmatrix}\cdot\begin{bmatrix}1 & -2\\3 & 1\end{bmatrix}^{-1}$

$$= \begin{bmatrix}1 & -2\\3 & 1\end{bmatrix}\begin{bmatrix}5 & 0\\0 & -2\end{bmatrix}\begin{bmatrix}\frac{1}{7} & \frac{2}{7}\\-\frac{3}{7} & \frac{1}{7}\end{bmatrix}$$

$$= \begin{bmatrix}5 & 4\\15 & -2\end{bmatrix}\begin{bmatrix}\frac{1}{7} & \frac{2}{7}\\-\frac{3}{7} & \frac{1}{7}\end{bmatrix}$$

$$= \begin{bmatrix}-1 & 2\\3 & 4\end{bmatrix}$$

$$= A$$

$A = PDP^{-1}$ is satisfied for $P = \begin{bmatrix}1 & -2\\3 & 1\end{bmatrix}$ $D = \begin{bmatrix}5 & 0\\0 & -2\end{bmatrix}$.

5. Proof: $A^n = PD^nP^{-1}$

base: $n=1$: $A = PDP^{-1}$, true

induction: $n=k+1$: $A^{k+1} = A \cdot A^k$

$$= \underbrace{(PDP^{-1})}_{\text{base case}} \cdot \underbrace{(PD^kP^{-1})}_{\text{assumption}}$$

$$= PDD^{k+1}P^{-1}$$

$$= PD^{k+1}P^{-1}$$, true

Conclusion:

Proved by induction: $A^n = PD^nP^{-1}$.

Hence $A^n = \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 5^n & 0 \\ 0 & -2^n \end{bmatrix} \begin{bmatrix} \frac{1}{7} & \frac{2}{7} \\ -\frac{3}{7} & \frac{1}{7} \end{bmatrix}$

$= \frac{1}{7} \begin{bmatrix} 1 & -2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5^n & 0 \\ 0 & -2^n \end{bmatrix}$.