Neural Networks – Exercise Sheet 1

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Exercise 1.1 – *Matrix Properties*

Given the matrix A with:

$$A = \begin{bmatrix} 4 & 2 & -1 \\ 2 & -1 & y \\ x & -2 & -8 \end{bmatrix}$$

a) *Compute values for x and y so that A*

i. *is symmetric*

Since it's known by definition that any symmetric matrix A has property $A = A^T$, we can assume following:

$$\begin{bmatrix} 4 & 2 & -1 \\ 2 & -1 & y \\ x & -2 & -8 \end{bmatrix} = \begin{bmatrix} 4 & 2 & x \\ 2 & -1 & -2 \\ -1 & y & -8 \end{bmatrix}$$

Thus, $x = -1, y = -2$

ii. *is an orthogonal matrix*

By the definition, orthogonal matrix is the matrix which satisfies the following condition: $AA^T = I$.

Thus

$$\begin{bmatrix} 4 & 2 & -1 \\ 2 & -1 & y \\ x & -2 & -8 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 & x \\ 2 & -1 & -2 \\ -1 & y & -8 \end{bmatrix} = I$$

After multiplying those two matrices we get following:

$$\begin{bmatrix} 4 \cdot 4 + 2 \cdot 2 + (-1) \cdot (-1) & 4 \cdot 2 + 2 \cdot (-1) + (-1) \cdot y & 4 \cdot x + 2 \cdot (-2) + (-1) \cdot (-8) \\ 2 \cdot 4 + (-1) \cdot 2 + y \cdot (-1) & 2 \cdot 2 + (-1) \cdot (-1) + y \cdot y & 2 \cdot x + (-1) \cdot (-2) + y \cdot (-8) \\ x \cdot x + (-2) \cdot (-2) + (-8) \cdot (-1) & x \cdot 2 + (-2) \cdot (-1) + (-8) \cdot y & x \cdot x + (-2) \cdot (-2) + (-8) \cdot (-8) \end{bmatrix}$$

Thus we have following resulting matrix which has to be by definition equal to identity matrix:

$$\begin{bmatrix} 21 & 6 - y & 4x + 4 \\ 6 - y & 5 + y^2 & 2x + 2 - 8y \\ 4x + 4 & 2x + 2 - 8y & x^2 + 68 \end{bmatrix} = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\Longrightarrow$ since square values of x and y are negative, it would not be possible to generate such values that given matrix would become orthogonal.

iii. *has rank 2*

Rank of the matrix is equal to the size of square submatrix that has nonzero determinant.
Hence, the rank of the singular matrix 3×3 will be equal two.
Knowing that, we can use the assumption made in *(iv)*, which means given matrix has the rank=2 if $x = 0$ and $y = -8.5$

iv. *is singular*

According to the definition, matrix is called singular if it doesn't have an inverse matrix, which means iff the determinant of such matrix is equal to 0.
Let's start by computing the equation for the determinant using the following:

$$detA = aei - afh - bdi + bfg + cdh - ceg$$

for the matrix given as:

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

$$detA = 4 \times (-1) \times (-8) - 4 \times y \times (-2) - 2 \times 2 \times (-8)$$
$$+2 \times y \times x + (-1) \times 2 \times (-2) - (-1) \times (-1) \times x = 68 - x + 2xy + 8y$$

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

After regrouping this equation, get following:

$$4(17 + 2y) + x(2y - 1) = 0$$

If we assume $x = 0$, then $y = -8.5$, thus the needed pair of values is: $x = 0$, and $y = -8.5$

We can perform fast check by finding eigenvalues for the resulting matrix, by following the rule, that one of eigenvalues of the singular matrix must be zero:

```
>> A=[4 2 -1
      2 -1 -8.5
      0 -2 -8];
>> eig(A)

ans =

    4.9330
    0.0000
   -9.9330
```

b) *Set x = 4, y = 0 and compute the Eigendecompostion of the resulting matrix.*

Given matrix:

$$\begin{bmatrix} 4 & 2 & -1 \\ 2 & -1 & 0 \\ 4 & -2 & -8 \end{bmatrix}$$

By performing Matlab functions, get following results:

```
>> A=[4 2 -1
      2 -1 0
      4 -2 -8]; %given Matrix
>> [V, D]=eig(A) %V - right eigenvectors; D - Matrix which has
eigenvalues as diagonal elements

V =
   -0.9119    0.0907    0.3537
   -0.3335   -0.0276   -0.7956
   -0.2391    0.9955    0.4919

D =
    4.4692         0         0
         0   -7.5800         0
         0         0   -1.8892
```

$\lambda_1 = 4.4692, v_1 = (-0.09119, -0.3335, -0.2391\ )$

$\lambda_2 = -7.5800, v_2 = (0.0907, -0.0276,\ 0.9955)$

$\lambda_3 = -1.8892, v_3 = (0.3537, -0.7956, 0.4919)$

Exercise 1.2 – Eigenvalues

a) *Find a matrix $B \in \mathbb{R}^{2x2}$ whose eigenvalues are 1 and 4 with the corresponding eigenvectors (3, 1) and (2, 1). Show your intermediate steps explicitly.*

$$\lambda_1 = 1 \qquad N_1 = (3, 1)$$
$$\lambda_2 = 4 \qquad N_2 = (2, 1)$$

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

$$V = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Since $B = V\Lambda V^{-1}$, then $\mathrm{B} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}\frac{1}{(3\times1)-(2\times1)}\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix}$

$$\mathrm{B} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \Rightarrow \mathrm{B} = \begin{bmatrix} 3 & 8 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix}$$

$$\Rightarrow \mathrm{B}_2 = \begin{bmatrix} -5 & 18 \\ -3 & 10 \end{bmatrix}$$

b) *Let A and B be two n × n matrices. Show that if λ is an eigenvalue of AB, then it is also an eigenvalue of BA.*
For AB and BA to have same eigenvalues means they must have same characteristic equation. Thus, for the solution it will be sufficient to show that AB and BA have the same characteristics polynomial.
Hence,

$$\begin{aligned} P_{AB} &= \det(AB - \lambda I) \\ &= \det(ABAA^{-1} - \lambda AA^{-1}) \\ &= \det[A(BA - \lambda I)A^{-1}] \\ &= \det(A)\det(BA - \lambda I)\frac{1}{\det(A)} \\ &= det\,(BA - \lambda I) = P_{BA} \end{aligned}$$

thus if λ is an eigenvalue of A, then it is also an eigenvalue of BA.

Exercise 1.3 – *Covariance Matrix*

In machine learning, we want to find and learn dependencies of features in given data. From the covariance of the considered features one can obtain how much two features behave similarly. This ability will reappear later in the lecture.

Let $X \in \mathbb{R}^{mxn}$ be an arbitrary matrix. The covariance matrix $C = X^TX$.

a) *Show that C is always a positive semidefinite matrix i.e. $v^TCv \geq 0$ for all non-zero $v \in \mathbb{R}^n$.*
The covariance matrix C is given as: X^TX. Therefore for all non-zero $v \in \mathbb{R}^n$ we can write following:

$$v^TCv = v^TX^TXv$$

Let $y = Xv$, $y^T = v^TX^T$.
Therefore, we can write:

$$v^TCv = y^Ty$$

From L_2 norm we know that:

$$y^Ty = \|y\|_2$$

and $\|y\|_2 \geq 0$
Therefore, $v^TCv \geq 0$ for all non-zero $v \in \mathbb{R}^n$.

b) *Let $U \in \mathbb{R}^{nxn}$ be an orthogonal matrix with $UU^T = U^TU = \mathbb{I}$, which is the identity matrix. Show that $\|U^Tx\|_2^2 = 1$ for any vector $X \in \mathbb{R}^n$ with $\|x\|_2 = 1$.*
Let's represent $\|U^Tx\|_2^2$ as following:

$$\|U^Tx\|_2^2 = x^TUU^Tx$$

We know that $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix with $UU^T = U^TU = \mathbb{I}$.
Therefore = x^Tx.
Thus, $\|U^Tx\|_2^2 = \|x\|_2$
Given that $\|x\|_2 = 1$ we can write that $\|U^Tx\|_2^2 = 1$.

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

c) *In the following, we want to minimize the term $v^T Cv$ with $\|v\|_2 = 1$. The Raleigh-Ritz principle allows us to solve this optimization problem using the Eigendecomposition of C.*
Prove that $min_{\|v\|_2=1} v^T Cv = \lambda_{min}$ where λ_{min} is the smallest eigenvalue of C. What can you say about v leading to the minimal value?
Since C is symmetric, we know that it has all real eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.
Therefore, $C = X^T \wedge X$ where $\wedge$ is diagonal and X is orthogonal.
So,

$$R(v) = v^T Cv = v^T X^T \wedge Xv = (X_v)^T \wedge X_v$$

Then, we can also represent it as:

$$R(X^T v) = (XX^T v)^T \wedge XX^T v = v^T \wedge v = \lambda_1 v_1^2 + \cdots + \lambda_n v_n^2$$

Now we can call the vector $w = X^T v$, then we have

$$R(w) = \lambda_1 v_1^2 + \cdots + \lambda_n v_n^2$$

We know that $\|v\|_2 = 1$, so
$R(w) = \lambda_1 v_1^2 + \cdots + \lambda_n v_n^2$ where $v_1^2 + \cdots + v_n^2 = 1$
Obviously, these values $R(w)$ lie in $[\lambda_1, \lambda_n]$, the minimum will lead to λ_1, which is the smallest eigenvalue.
Thus,

$$min_{\|v\|_2=1} v^T Cv = \lambda_1$$