lecture 7

A-exercises – Solutions

Exercise A1: Matrix operations

Given

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}.$$

1. Addition is element-wise:

$$A + B = \begin{pmatrix} 1+2 & 2+0 \\ 3+1 & 4+3 \end{pmatrix} = \boxed{\begin{pmatrix} 3 & 2 \\ 4 & 7 \end{pmatrix}}.$$

2. For AB, multiply rows of A by columns of B:

$$AB = \begin{pmatrix} 1\cdot 2 + 2\cdot 1 & 1\cdot 0 + 2\cdot 3 \\ 3\cdot 2 + 4\cdot 1 & 3\cdot 0 + 4\cdot 3 \end{pmatrix} = \boxed{\begin{pmatrix} 4 & 6 \\ 10 & 12 \end{pmatrix}}.$$

3. Similarly,

$$BA = \begin{pmatrix} 2\cdot 1 + 0\cdot 3 & 2\cdot 2 + 0\cdot 4 \\ 1\cdot 1 + 3\cdot 3 & 1\cdot 2 + 3\cdot 4 \end{pmatrix} = \boxed{\begin{pmatrix} 2 & 4 \\ 10 & 14 \end{pmatrix}}.$$

Since $AB \neq BA$, matrix multiplication is in general **not commutative**. Thus, in general,

$$AB \neq BA.$$

Exercise A2: Linear system

We have

$$\begin{cases} 2x + y = 5, \\ x - y = 1. \end{cases}$$

Adding the two equations gives

$$3x = 6, \qquad x = 2.$$

Substitution into $x - y = 1$ gives

$$2 - y = 1, \qquad y = 1.$$

Hence

$$\boxed{x = 2, \qquad y = 1}.$$

In matrix form,

$$\boxed{\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}}$$

so that

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$

Exercise A3: Least squares

Given Data We are given four data points (x_i, y_i) for $i = 1, 2, 3, 4$:

Index (i)	x	y
1	1	2.1
2	2	3.9
3	3	6.2
4	4	8.1

We fit the linear model:

$$y = ax + b$$

where $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$ represents the vector of unknown parameters.

1. Construction of the Design Matrix X and Vector y

The linear system in matrix form is written as $X\theta \approx y$. The design matrix X contains the x-values in the first column and a column of ones for the intercept b in the second column:

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 2.1 \\ 3.9 \\ 6.2 \\ 8.1 \end{pmatrix}$$

2. Determination of Least Squares Parameters

The optimal parameters $\hat{\theta} = \begin{pmatrix} a \\ b \end{pmatrix}$ are given by the normal equation:

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

Step 2.1: Compute $X^T X$

$$X^T X = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 1+4+9+16 & 1+2+3+4 \\ 1+2+3+4 & 1+1+1+1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

Step 2.2: Compute $(X^T X)^{-1}$

The determinant of $X^T X$ is $\det(X^T X) = (30 \times 4) - (10 \times 10) = 120 - 100 = 20$.

$$(X^T X)^{-1} = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} = \begin{pmatrix} 0.2 & -0.5 \\ -0.5 & 1.5 \end{pmatrix}$$

Step 2.3: Compute $X^T y$

$$X^T y = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2.1 \\ 3.9 \\ 6.2 \\ 8.1 \end{pmatrix} = \begin{pmatrix} (1 \cdot 2.1) + (2 \cdot 3.9) + (3 \cdot 6.2) + (4 \cdot 8.1) \\ 2.1 + 3.9 + 6.2 + 8.1 \end{pmatrix} = \begin{pmatrix} 60.9 \\ 20.3 \end{pmatrix}$$

Step 2.4: Calculate Parameter Vector $\hat{\theta}$

$$\hat{\theta} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0.2 & -0.5 \\ -0.5 & 1.5 \end{pmatrix} \begin{pmatrix} 60.9 \\ 20.3 \end{pmatrix}$$

$$a = (0.2 \cdot 60.9) + (-0.5 \cdot 20.3) = 12.18 - 10.15 = 2.03$$

$$b = (-0.5 \cdot 60.9) + (1.5 \cdot 20.3) = -30.45 + 30.45 = 0.00$$

Conclusion

The least squares fit yields:

- **Slope (a):** 2.03
- **Intercept (b):** 0.00

The fitted line equation is $y = 2.03x$.

Exercise A4: Beer–Lambert calibration

The calibration model is

$$A = mc + b.$$

1. The parameter m is the **slope**: it describes how strongly the measured absorbance A changes when the concentration c changes. The parameter b is the **intercept**: it represents the predicted absorbance at $c = 0$.

2. Given measured calibration pairs (c_i, A_i), least squares chooses m and b to minimize

   $$\sum_i \left(A_i - (mc_i + b)\right)^2 .$$

   In matrix form,

   $$\mathbf{A} = X \begin{pmatrix} m \\ b \end{pmatrix}, \qquad X = \begin{pmatrix} c_1 & 1 \\ \vdots & \vdots \\ c_N & 1 \end{pmatrix}.$$

   The least-squares parameters satisfy the normal equations

   $$X^T X \boldsymbol{\beta} = X^T \mathbf{A}, \qquad \boldsymbol{\beta} = \begin{pmatrix} m \\ b \end{pmatrix}.$$

3. Once m and b are known, an unknown concentration can be estimated from a measured absorbance A by solving

   $$A = mc + b.$$

   Hence

   $$\boxed{c = \frac{A - b}{m}}$$

   provided $m \neq 0$.

Exercise A5: Activation functions

For

$$z = \begin{pmatrix} -2 \\ 0.5 \\ 3 \end{pmatrix},$$

the ReLU function is applied component-wise:

$$\phi(z) = \max(0, z).$$

Therefore

$$\boxed{\phi(z) = \begin{pmatrix} 0 \\ 0.5 \\ 3 \end{pmatrix}}.$$

ReLU introduces nonlinearity because it is not a linear function. In particular,

$$\phi(z) = \begin{cases} 0, & z < 0, \\ z, & z \geq 0. \end{cases}$$

Without an activation function, a composition of linear transformations would still be a linear transformation. The nonlinear activation therefore allows a neural network to represent nonlinear relationships.

Exercise A6: AI model interpretation

The neural-network model is

$$\hat{y} = W_2\phi(W_1x + b_1) + b_2.$$

1. W_1 and W_2 are **weight matrices**. They determine how the input information is linearly transformed between layers. During training, these weights are adjusted to fit the data.
2. b_1 and b_2 are **bias vectors**. They shift the outputs of the corresponding linear transformations and give the network additional flexibility.
3. ϕ is the **activation function**, such as ReLU. It introduces nonlinearity between the linear layers, allowing the network to represent nonlinear relationships.

The model can therefore be viewed as a sequence:

$$x \longrightarrow W_1x + b_1 \longrightarrow \phi(W_1x + b_1) \longrightarrow W_2\phi(W_1x + b_1) + b_2 \longrightarrow \hat{y}.$$

Final conclusion

The central message is

$$\boxed{\text{Artificial intelligence is built on mathematical foundations.}}$$

Matrices describe information and linear transformations. Least squares provides a systematic way to learn parameters from data. Activation functions introduce the nonlinearity needed for expressive neural-network models. Together, these concepts form an important mathematical basis for modern data analysis in science and engineering.