Appendix D Introduction to Linear Algebra

John D. Kelleher and Brian Mac Namee and Aoife D'Arcy

- A **scalar** is a single number.
- A **matrix** is a 2-dimensional ($n \times m$) array of numbers.

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \dots & \dots & \dots & \dots \\ c_{n,1} & c_{1,2} & \dots & c_{n,m} \end{bmatrix}$$

- Each element in a matrix is identified by two indices, the row index and then the column index. For example,

$$\mathbf{C}[2,2] = c_{2,2}$$

- A **vector** is an array of numbers, organized in a specific order.
- A vector can be either a column vector or a row vector.
- For example, **a** is a column vector, and **b** is a row vector.

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 & \ldots & b_m \end{bmatrix}$$

- Each element in a vector is identified by a single index. For example:

$$\mathbf{a}[2] = a_2$$
$$\mathbf{b}[2] = b_2$$

- Vectors are often treated as special cases of matrices:
  - a column vector can be thought of as a matrix with just one column,
  - a row vector can be thought of as a matrix with a single row.

- The transpose of a vector converts a column vector to a row vector, and vice versa.
- If $\mathbf{a}$ is a vector, then we write $\mathbf{a}^\intercal$ for the transpose of $\mathbf{a}$.

Example

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}$$

$$\mathbf{a}^\intercal = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$$

- The transpose of a matrix flips the matrix on its main diagonal
  - the main diagonal of a matrix contain all the elements whose indices are equal, e.g., $c_{1,1}$, $c_{2,2}$, and so on.
- To create the transpose of a matrix, take the first row of the matrix and write it as the first column; then write the second row of the matrix and write it as the second column; and so on.

Example

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} \\ c_{3,1} & c_{3,2} & c_{3,3} & c_{3,4} \end{bmatrix}$$

$$\mathbf{C}^{\mathsf{T}} = \begin{bmatrix} c_{1,1} & c_{2,1} & c_{3,1} \\ c_{1,2} & c_{2,2} & c_{3,2} \\ c_{1,3} & c_{2,3} & c_{3,3} \\ c_{1,4} & c_{2,4} & c_{3,4} \end{bmatrix}$$

- In general, there is no special symbol used to denote a matrix product.
- We write the matrix product by writing the names of the two matrices side by side.

Example

For example, $\mathbf{DE}$ is the way we write the product for two matrices $\mathbf{D}$ and $\mathbf{E}$, although sometimes a *dot* may be inserted between the two matrices (a $\cdot$ is frequently used to highlight that one or both of the matrices is a vector):

$$\mathbf{DE} = \mathbf{D} \cdot \mathbf{E}$$

- To multiply one matrix by another, the number of columns in the matrix on the left of the product must equal the number of rows in the matrix on the right.
- If this condition does not hold, then the product of the two matrices is not defined.

Example

- if **D** is a 2×3 matrix (i.e., a matrix with 2 rows and 3 columns) and **E** is a 3×3 matrix, then
    - the matrix product **DE** is defined, because the number of columns in **D** (3) equals the number of rows in **E** (3).
    - the matrix product **ED** is not defined because the number of columns in **E** (3) is not equal to the number of rows in **D** (2).
    - the matrix product $\mathbf{ED}^{\mathsf{T}}$ is defined because $\mathbf{D}^{\mathsf{T}}$ is a 3×2 matrix and the number of columns is **E** (3) equals the number of rows in $\mathbf{D}^{\mathsf{T}}$ (3).

- The result of multiplying two matrices is another matrix whose dimensions are equal to the number of rows in the left matrix and the number of columns in the right matrix.
- Multiplying a 2×3 matrix by a 3×3 matrix results in a 2×3.
- Each value in the resulting matrix is calculated as follows, where i iterates over the columns in the first matrix (**D**) and the rows in the second matrix (**E**)

$$\mathbf{DE}_{r,c} = \sum_{i} \mathbf{D}[r, i] \times \mathbf{E}[i, c]$$

Example

$$\mathbf{D} = \begin{bmatrix} d_{1,1} & d_{1,2} & d_{1,3} \\ d_{2,1} & d_{2,2} & d_{2,3} \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} \\ e_{2,1} & e_{2,2} & e_{2,3} \\ e_{3,1} & e_{3,2} & e_{3,3} \end{bmatrix}$$

$\mathbf{DE} =$

$$\begin{bmatrix} (d_{1,1}e_{1,1})+(d_{1,2}e_{2,1})+(d_{1,3}e_{3,1}) & (d_{1,1}e_{1,2})+(d_{1,2}e_{2,2})+(d_{1,3}e_{3,2}) & (d_{1,1}e_{1,3})+(d_{1,2}e_{2,3})+(d_{1,3}e_{3,3}) \\ (d_{2,1}e_{1,1})+(d_{2,2}e_{2,1})+(d_{2,3}e_{3,1}) & (d_{2,1}e_{1,2})+(d_{2,2}e_{2,2})+(d_{2,3}e_{3,2}) & (d_{2,1}e_{1,3})+(d_{2,2}e_{2,3})+(d_{2,3}e_{3,3}) \end{bmatrix}$$

- The product of two vectors of the same dimensions is known as the **dot product**.

Example

- if **F** and **G** are row vectors which have dimensions 1×3

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} g_1 & g_2 & g_3 \end{bmatrix}$$

- the **dot product** of **F** and **G**, written $\mathbf{F} \cdot \mathbf{G}$ is equivalent to the matrix product $\mathbf{F}\mathbf{G}^\mathsf{T}$

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \end{bmatrix} \quad \mathbf{G}^\mathsf{T} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

$$\mathbf{F} \cdot \mathbf{G} = (f_1 g_1) + (f_2 g_2) + (f_3 g_3)$$

- In **deep learning** we frequently use an elementwise product of two matrices, known as the **Hadamard product**.
- The symbol $\odot$ denotes the Hadamard product, and the Hadamard product of two matrices $\mathbf{D}$ and $\mathbf{E}$ is written $\mathbf{D} \odot \mathbf{E}$.
- The Hadamard product assumes that both matrices have the same dimensions, and it produces a matrix with the same dimensions as the two inputs.
- Each value in the resulting matrix is the product of the corresponding cells in the two input matrices:

$$\mathbf{DE}_{r,c} = \mathbf{D}[r,c] \times \mathbf{E}[r,c]$$

Example

$$\mathbf{D} = \begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{bmatrix}$$

$$\mathbf{D} \odot \mathbf{E} = \begin{bmatrix} (d_{1,1}e_{1,1}) & (d_{1,2}e_{1,2}) \\ (d_{2,1}e_{2,1}) & (d_{2,2}e_{2,2}) \end{bmatrix}$$

Summary