Deep Learning Principles & Applications

Chapter 1 – Linear Algebra for Deep Learning

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

Vectors from a Data Matrix

Vectors from a Data Matrix

Data from 4 patients and 3 features:

Vectors from a Data Matrix

Data from 4 patients and 3 features:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

Vectors from a Data Matrix

Data from 4 patients and 3 features:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

A vector is a columnwise listing of numbers.

Vectors from a Data Matrix

Data from 4 patients and 3 features:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

A vector is a columnwise listing of numbers.

1st feature vector (heart rate) for all patients:

$$\mathbf{x_1} = \begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}$$

Vectors from a Data Matrix

Data from 4 patients and 3 features:

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

A vector is a columnwise listing of numbers.

1st feature vector (heart rate) for all patients:

$$\mathbf{x_1} = \begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}$$

1st patient vector for all features:

$$\mathbf{x}^{(1)} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$$

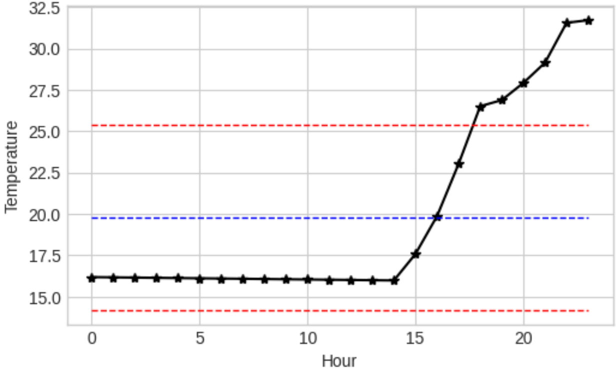
Vector Example

Vector Example

Hourly temperature ($^\circ$C) in downtown San Francisco on October 2, 2012:

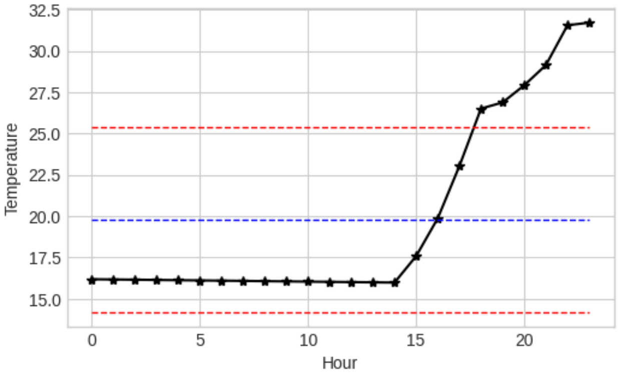
Vector Example

Hourly temperature (°C) in downtown San Francisco on October 2, 2012:



Vector Example

Hourly temperature (°C) in downtown San Francisco on October 2, 2012:



$$\mathbf{t} = \begin{bmatrix} 16.18 \\ 16.17 \\ 16.15 \\ 16.14 \\ 16.12 \\ 16.11 \\ 16.09 \\ 16.08 \\ 16.07 \\ 16.05 \\ 16.04 \\ 16.02 \\ 16.01 \\ 15.99 \\ 15.98 \\ 17.58 \\ 19.87 \\ 23.03 \\ 26.51 \\ 26.88 \\ 27.92 \\ 29.14 \\ 31.55 \\ 31.71 \end{bmatrix}$$

Vector Example-continued

Vector Example-continued

Consider the count of words in the following sentence:

Vector Example-continued

Consider the count of words in the following sentence:

Word count vectors are used in computer based document analysis. Each entry of the word count vector is the number of times the associated dictionary word appears in the document.

Vector Example-continued

Consider the count of words in the following sentence:

Word count vectors are used in computer based document analysis. Each entry of the word count vector is the number of times the associated dictionary word appears in the document.

$$\begin{array}{l} \text{word} \\ \text{in} \\ \text{number} \\ \text{horse} \\ \text{the} \\ \text{document} \end{array} \qquad \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \\ 4 \\ 2 \end{bmatrix}$$

Attributes of a Vector

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.
- A vector of size n is called an n-vector.

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.
- A vector of size n is called an n-vector.
- The ith element of the vector $\mathbf{x}$ is denoted as x_i.

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.
- A vector of size n is called an n-vector.
- The ith element of the vector $\mathbf{x}$ is denoted as x_i.
- For example, from the patient data matrix:

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.
- A vector of size n is called an n-vector.
- The ith element of the vector $\mathbf{x}$ is denoted as x_i.
- For example, from the patient data matrix: $\mathbf{x}^{(1)}$ is a 3-vector representing the 1st patient:

Attributes of a Vector

- The elements (or entries or coefficients or components) of a vector are the values that comprise it.
- The size (also called the dimension) of a vector is the number of elements it contains.
- A vector of size n is called an n-vector.
- The ith element of the vector $\mathbf{x}$ is denoted as x_i.
- For example, from the patient data matrix: $\mathbf{x}^{(1)}$ is a 3-vector representing the 1st patient:

$$\mathbf{x}^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix} = \underbrace{\begin{bmatrix} \text{HR in BPM} \\ \text{BP in mm Hg} \\ \text{Temp in } ^\circ\text{C} \end{bmatrix}}_{\text{for patient-1}}.$$

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$(-2)\begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

If $\mathbf{x}$ and $\mathbf{y}$ are word count vectors (using the same dictionary) for two documents:

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

If $\mathbf{x}$ and $\mathbf{y}$ are word count vectors (using the same dictionary) for two documents:

- $\mathbf{x} + \mathbf{y}$ is the word count vector of a new document created by combining the original two (in either order);

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

If $\mathbf{x}$ and $\mathbf{y}$ are word count vectors (using the same dictionary) for two documents:

- $\mathbf{x} + \mathbf{y}$ is the word count vector of a new document created by combining the original two (in either order);
- $\mathbf{x} - \mathbf{y}$ gives the number of times more each word appears in the first document than the second.

Vector Operations: Addition, Subtraction, Scalar Multiplication

Vector addition, subtraction, and scalar multiplication are performed elementwise:

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}$$

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

If $\mathbf{x}$ and $\mathbf{y}$ are word count vectors (using the same dictionary) for two documents:

- $\mathbf{x} + \mathbf{y}$ is the word count vector of a new document created by combining the original two (in either order);
- $\mathbf{x} - \mathbf{y}$ gives the number of times more each word appears in the first document than the second.

If $\mathbf{z}$ is a vector representing an audio signal (sound pressure), the scalar-vector product $2\mathbf{z}$ is the same audio signal perceived twice as loud.

Matrices

Matrices

A matrix is a tabular arrangement of numbers written between rectangular brackets:

Matrices

A matrix is a tabular arrangement of numbers written between rectangular brackets:

4×3-Patient data matrix $\mathbf{X}$

Matrices

A matrix is a tabular arrangement of numbers written between rectangular brackets:

4×3-Patient data matrix $\mathbf{X}$

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

Matrices

A matrix is a tabular arrangement of numbers written between rectangular brackets:

4×3-Patient data matrix $\mathbf{X}$

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

8×8 Grayscale image matrix $\mathbf{A}$

Matrices

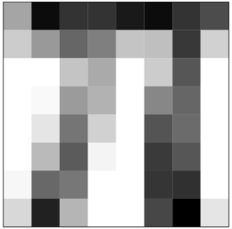
A matrix is a tabular arrangement of numbers written between rectangular brackets:

4×3-Patient data matrix $\mathbf{X}$

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

8×8 Grayscale image matrix $\mathbf{A}$

0.65 0.05 0.20



0.28 0.00 0.90

$$\begin{bmatrix} 0.65 & 0.05 & 0.2 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0.28 & 0.0 & 0.9 \end{bmatrix}$$

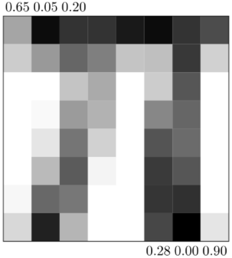
Matrices

A matrix is a tabular arrangement of numbers written between rectangular brackets:

4×3-Patient data matrix $\mathbf{X}$

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

8×8 Grayscale image matrix $\mathbf{A}$



$$\begin{bmatrix} 0.65 & 0.05 & 0.2 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0.28 & 0.0 & 0.9 \end{bmatrix}$$

The 4-row and 3-column patient data matrix $\mathbf{X}$ can be seen as a 3-vector (corresponding to each patient) repeating 4 times.

Attributes of a Matrix

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns:

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix:

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix: for example, the patient data matrix $\mathbf{X}$ is a 4×3-matrix.

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix: for example, the patient data matrix $\mathbf{X}$ is a 4×3-matrix.
- The elements (or entries or coefficients) of a matrix are the values that comprise it.

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix: for example, the patient data matrix $\mathbf{X}$ is a 4×3-matrix.
- The elements (or entries or coefficients) of a matrix are the values that comprise it.
- The notation x_{ij} specifies the element of the matrix $\mathbf{X}$ in the ith row and jth column:

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix: for example, the patient data matrix $\mathbf{X}$ is a 4×3-matrix.
- The elements (or entries or coefficients) of a matrix are the values that comprise it.
- The notation x_{ij} specifies the element of the matrix $\mathbf{X}$ in the ith row and jth column: for example, for the patient data matrix, x_{12} represents the element in the 1st row and 2nd column,

Attributes of a Matrix

- The size (also called the dimension) of a matrix is the number of rows and columns: for example, the patient data matrix $\mathbf{X}$ has 4 rows (patients) and 3 columns (features).
- A matrix with m rows and n columns is denote as an $m \times n$-matrix: for example, the patient data matrix $\mathbf{X}$ is a 4×3-matrix.
- The elements (or entries or coefficients) of a matrix are the values that comprise it.
- The notation x_{ij} specifies the element of the matrix $\mathbf{X}$ in the ith row and jth column: for example, for the patient data matrix, x_{12} represents the element in the 1st row and 2nd column, which is also the 2nd feature (BP) of the 1st patient.

Matrix Operations

Matrix Operations

- Matrix operations, just like vectors, are performed elementwise:

Matrix Operations

- Matrix operations, just like vectors, are performed elementwise:

$$\underbrace{\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}}_{\text{Matrix addition}} \qquad \underbrace{(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}}_{\text{Matrix-scalar multiplication}}$$

Matrix Operations

- Matrix operations, just like vectors, are performed elementwise:

$$\underbrace{\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}}_{\text{Matrix addition}} \quad \underbrace{(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}}_{\text{Matrix-scalar multiplication}}$$

- The transpose of a matrix $\mathbf{X}$ denoted as $\mathbf{X}^{\mathrm{T}}$ is the matrix obtained by switching the rows and columns of $\mathbf{X}$:

Matrix Operations

- Matrix operations, just like vectors, are performed elementwise:

$$\underbrace{\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}}_{\text{Matrix addition}} \quad \underbrace{(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}}_{\text{Matrix-scalar multiplication}}$$

- The transpose of a matrix $\mathbf{X}$ denoted as $\mathbf{X}^{\mathrm{T}}$ is the matrix obtained by switching the rows and columns of $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 76 & 126 & 38.0 \\ 74 & 120 & 38.0 \\ 72 & 118 & 37.5 \\ 78 & 136 & 37.0 \end{bmatrix} \Rightarrow \mathbf{X}^{\mathrm{T}} = \begin{bmatrix} 76 & 74 & 72 & 78 \\ 126 & 120 & 118 & 136 \\ 38.0 & 38.0 & 37.5 & 37.0 \end{bmatrix}$$

Different Views of a Matrix

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(\mathbf{1})} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that ${\mathbf{x}^{(\mathbf{1})}}^{\mathrm{T}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(1)} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that ${\mathbf{x}^{(1)}}^{\mathrm{T}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.
- A matrix can be represented in terms of its columns or rows:

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(1)} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that ${\mathbf{x}^{(1)}}^{\mathrm{T}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.
- A matrix can be represented in terms of its columns or rows:

$$\mathbf{X} = \begin{bmatrix} \underbrace{\begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}}_{\text{column-1}} & \underbrace{\begin{bmatrix} 126 \\ 120 \\ 118 \\ 136 \end{bmatrix}}_{\text{column-2}} & \underbrace{\begin{bmatrix} 38.0 \\ 38.0 \\ 37.5 \\ 37.0 \end{bmatrix}}_{\text{column-3}} \end{bmatrix}$$

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(\mathbf{1})} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that ${\mathbf{x}^{(\mathbf{1})}}^{\mathrm{T}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.
- A matrix can be represented in terms of its columns or rows:

$$\mathbf{X} = \left[\underbrace{\begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}}_{\text{column-1}} \underbrace{\begin{bmatrix} 126 \\ 120 \\ 118 \\ 136 \end{bmatrix}}_{\text{column-2}} \underbrace{\begin{bmatrix} 38.0 \\ 38.0 \\ 37.5 \\ 37.0 \end{bmatrix}}_{\text{column-3}} \right]$$

$$= \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} \end{bmatrix}$$

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(1)} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that $\mathbf{x}^{(1)^{\mathrm{T}}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.
- A matrix can be represented in terms of its columns or rows:

$$\mathbf{X} = \begin{bmatrix} \underbrace{\begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}}_{\text{column-1}} & \underbrace{\begin{bmatrix} 126 \\ 120 \\ 118 \\ 136 \end{bmatrix}}_{\text{column-2}} & \underbrace{\begin{bmatrix} 38.0 \\ 38.0 \\ 37.5 \\ 37.0 \end{bmatrix}}_{\text{column-3}} \end{bmatrix} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \underbrace{\begin{bmatrix} 76 & 126 & 38.0 \end{bmatrix}}_{\text{Row-1}} \\ \underbrace{\begin{bmatrix} 74 & 120 & 38.0 \end{bmatrix}}_{\text{Row-2}} \\ \underbrace{\begin{bmatrix} 72 & 118 & 37.5 \end{bmatrix}}_{\text{Row-3}} \\ \underbrace{\begin{bmatrix} 78 & 136 & 37.0 \end{bmatrix}}_{\text{Row-4}} \end{bmatrix}$$

Different Views of a Matrix

- A vector, such as $\mathbf{x}^{(1)} = \begin{bmatrix} 76 \\ 126 \\ 38 \end{bmatrix}$ can be viewed as a 3×1-matrix such that $\mathbf{x}^{(1)^{\mathrm{T}}} = \begin{bmatrix} 76 & 126 & 38 \end{bmatrix}$ is a 1×3-matrix.
- A matrix can be represented in terms of its columns or rows:

$$\mathbf{X} = \begin{bmatrix} \underbrace{\begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix}}_{\text{column-1}} & \underbrace{\begin{bmatrix} 126 \\ 120 \\ 118 \\ 136 \end{bmatrix}}_{\text{column-2}} & \underbrace{\begin{bmatrix} 38.0 \\ 38.0 \\ 37.5 \\ 37.0 \end{bmatrix}}_{\text{column-3}} \end{bmatrix} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \underbrace{\begin{bmatrix} 76 & 126 & 38.0 \end{bmatrix}}_{\text{Row-1}} \\ \underbrace{\begin{bmatrix} 74 & 120 & 38.0 \end{bmatrix}}_{\text{Row-2}} \\ \underbrace{\begin{bmatrix} 72 & 118 & 37.5 \end{bmatrix}}_{\text{Row-3}} \\ \underbrace{\begin{bmatrix} 78 & 136 & 37.0 \end{bmatrix}}_{\text{Row-4}} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^{(1)^{\mathrm{T}}} \\ \mathbf{x}^{(2)^{\mathrm{T}}} \\ \mathbf{x}^{(3)^{\mathrm{T}}} \\ \mathbf{x}^{(4)^{\mathrm{T}}} \end{bmatrix}$$

Tensors

Tensors

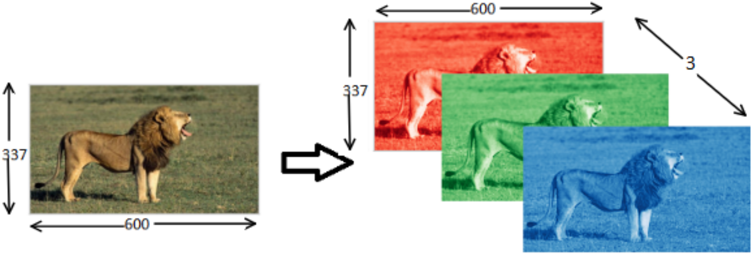
- A tensor is typically a 3-dimensional (or more) arrangement of numbers;

Tensors

- A tensor is typically a 3-dimensional (or more) arrangement of numbers; for example, 3-channel (RGB) color images can be represented as a 3D tensor

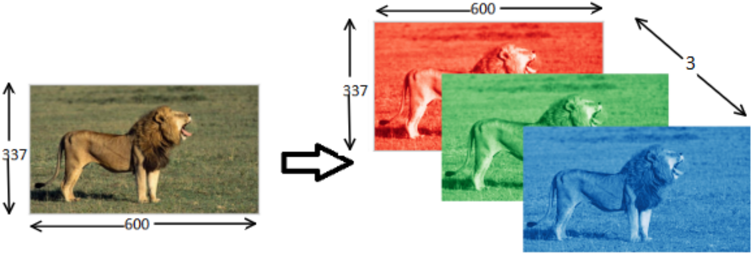
Tensors

- A tensor is typically a 3-dimensional (or more) arrangement of numbers; for example, 3-channel (RGB) color images can be represented as a 3D tensor



Tensors

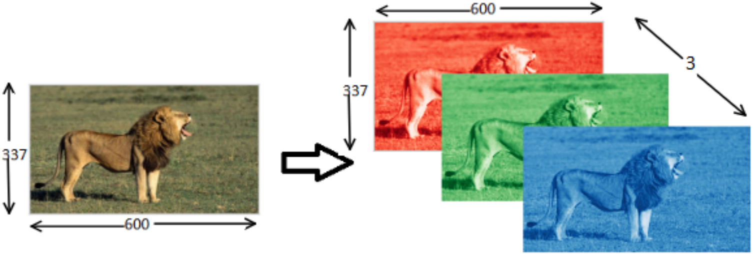
- A tensor is typically a 3-dimensional (or more) arrangement of numbers; for example, 3-channel (RGB) color images can be represented as a 3D tensor



- The image tensor has the shape $337 \times 600 \times 3$,

Tensors

- A tensor is typically a 3-dimensional (or more) arrangement of numbers; for example, 3-channel (RGB) color images can be represented as a 3D tensor



- The image tensor has the shape $337 \times 600 \times 3$, which can be seen as a 600×3-matrix (The R-G-B values for the horizontal 600 pixels from the top) repeating 337 times (for the vertical pixels).

Tensors–continued

Tensors–continued

- Here is another example of a 3D tensor incorporating time:

Tensors–continued

- Here is another example of a 3D tensor incorporating time: suppose we measured for 3 patients at 4 hourly timestamps 2 features (heart rate and blood pressure).

Tensors–continued

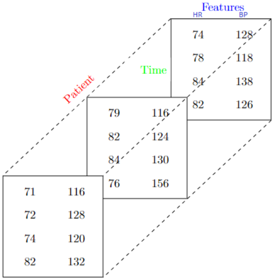
- Here is another example of a 3D tensor incorporating time: suppose we measured for 3 patients at 4 hourly timestamps 2 features (heart rate and blood pressure).
- The resulting 3D tensor could have the shape structure patients $\times$ timestamps $\times$ features;

Tensors–continued

- Here is another example of a 3D tensor incorporating time: suppose we measured for 3 patients at 4 hourly timestamps 2 features (heart rate and blood pressure).
- The resulting 3D tensor could have the shape structure patients $\times$ timestamps $\times$ features; an example tensor representing the data with shape $3 \times 4 \times 2$ is shown below:

Tensors–continued

- Here is another example of a 3D tensor incorporating time: suppose we measured for 3 patients at 4 hourly timestamps 2 features (heart rate and blood pressure).
- The resulting 3D tensor could have the shape structure patients $\times$ timestamps $\times$ features; an example tensor representing the data with shape $3 \times 4 \times 2$ is shown below:



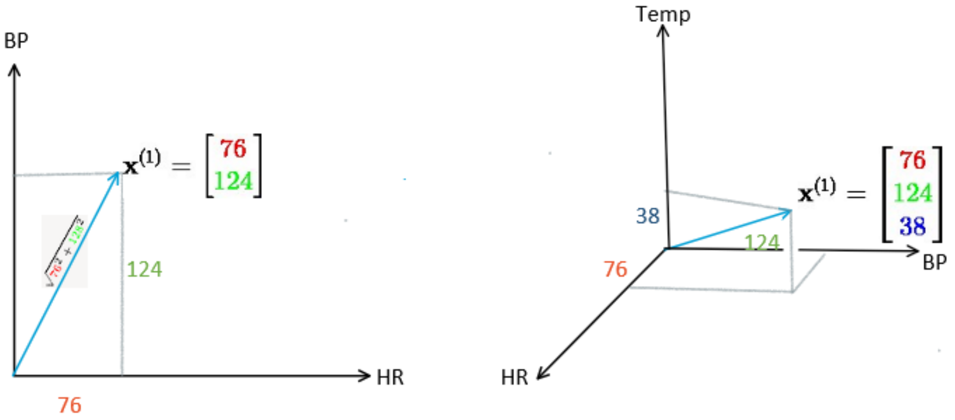
Geometric Representation of a Vector

Geometric Representation of a Vector

Vectors with 2 and 3 components can be easily visualized geometrically:

Geometric Representation of a Vector

Vectors with 2 and 3 components can be easily visualized geometrically:



Norm of a Vector

Norm of a Vector

- There are several ways to measure the geometric length of vector.

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties;

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.
- Such length-of-a-vector measurement definitions are called norms.

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.
- Such length-of-a-vector measurement definitions are called norms.
- Two popular norms in the deep learning-context are the l_1 and l_2 norms demonstrated using the following example:

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.
- Such length-of-a-vector measurement definitions are called norms.
- Two popular norms in the deep learning-context are the l_1 and l_2 norms demonstrated using the following example:

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \Rightarrow$$

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.
- Such length-of-a-vector measurement definitions are called norms.
- Two popular norms in the deep learning-context are the l_1 and l_2 norms demonstrated using the following example:

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \Rightarrow \underbrace{\|\mathbf{x}\|_1 = |1| + |-2| + |3|}_{l_1 \text{ norm}},$$

Norm of a Vector

- There are several ways to measure the geometric length of vector.
- Such measurement definitions should satisfy certain intuitive properties; such as, for example, when a vector is multiplied by a scalar, the resulting vector should have a length that is equal to the scaled (by the magnitude of the scalar) length of the original vector.
- Such length-of-a-vector measurement definitions are called norms.
- Two popular norms in the deep learning-context are the l_1 and l_2 norms demonstrated using the following example:

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \Rightarrow \underbrace{\|\mathbf{x}\|_1 = |1| + |-2| + |3|}_{l_1 \text{ norm}}, \underbrace{\|\mathbf{x}\|_2 = \sqrt{1^2 + (-2)^2 + 3^2}}_{l_2 \text{ norm}}$$

Dot Product

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$$

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$$

- The $\cdot$ (dot) symbol represents the computation of the dot product.

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n,$$

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$$

- The $\cdot$ (dot) symbol represents the computation of the dot product.
- An important identity is $\|\mathbf{x}\|_2^2 = \mathbf{x}\cdot\mathbf{x}$;

Dot Product

- The dot product of two vectors of same size is an operation that returns a scalar value through an elementwise multiplication and addition:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$$

- The $\cdot$ (dot) symbol represents the computation of the dot product.
- An important identity is $\|\mathrm{x}\|_2^2 = \mathrm{x} \cdot \mathrm{x}$; the square of l_2 norm of a vector is equal to the dot product of the vector with itself.

Dot Product Examples

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods,

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods, then the dot product $\mathbf{p} \cdot \mathbf{q}$ is the total cost of the goods.

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods, then the dot product $\mathbf{p} \cdot \mathbf{q}$ is the total cost of the goods.
- If $\mathbf{x}$ is an n-vector of the number of word occurrences in a document from a dictionary of n words and $\mathbf{w}$ is an n-vector with elements from $\{-1, 0, 1\}$,

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods, then the dot product $\mathbf{p} \cdot \mathbf{q}$ is the total cost of the goods.
- If $\mathbf{x}$ is an n-vector of the number of word occurrences in a document from a dictionary of n words and $\mathbf{w}$ is an n-vector with elements from $\{-1, 0, 1\}$, then the dot product $\mathbf{w} \cdot \mathbf{x}$ gives a (crude) measure of the sentiment (positive, negative, neutral) of that document.

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods, then the dot product $\mathbf{p} \cdot \mathbf{q}$ is the total cost of the goods.
- If $\mathbf{x}$ is an n-vector of the number of word occurrences in a document from a dictionary of n words and $\mathbf{w}$ is an n-vector with elements from $\{-1, 0, 1\}$, then the dot product $\mathbf{w} \cdot \mathbf{x}$ gives a (crude) measure of the sentiment (positive, negative, neutral) of that document.
- If $\mathbf{x}^{(1)}$ represents the 1st patient vector with HR, BP, and Temp values, and $\mathbf{w}$ is a 3-vector of weights applied to the 3 features,

Dot Product Examples

- If $\mathbf{p}$ represents a vector of prices of n goods, and $\mathbf{q}$ is a vector of quantities of the n goods, then the dot product $\mathbf{p} \cdot \mathbf{q}$ is the total cost of the goods.
- If $\mathbf{x}$ is an n-vector of the number of word occurrences in a document from a dictionary of n words and $\mathbf{w}$ is an n-vector with elements from $\{-1, 0, 1\}$, then the dot product $\mathbf{w} \cdot \mathbf{x}$ gives a (crude) measure of the sentiment (positive, negative, neutral) of that document.
- If $\mathbf{x}^{(1)}$ represents the 1st patient vector with HR, BP, and Temp values, and $\mathbf{w}$ is a 3-vector of weights applied to the 3 features, then the dot product $\mathbf{w} \cdot \mathbf{x}^{(1)}$ gives a weighted raw score for the 1st sample.

Angle between Vectors

Angle between Vectors

- Cauchy-Schwarz inequality:

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that
$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that
$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that

$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$$

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that
$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \Rightarrow \angle(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right).$$

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that

$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \Rightarrow \angle(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right).$$

- The angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ vary between 0 and π radians.

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that

$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \Rightarrow \angle(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right).$$

- The angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ vary between 0 and π radians.
- Note that when the dot product $\mathbf{x} \cdot \mathbf{y}$ is 0,

Angle between Vectors

- Cauchy-Schwarz inequality: for any two n-vectors $\mathbf{x}, \mathbf{y}$, it is always true that
$$-1 \leq \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \leq 1.$$

- This is used to define the cosine of the angle and the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \Rightarrow \angle(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right).$$

- The angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ vary between 0 and π radians.
- Note that when the dot product $\mathbf{x} \cdot \mathbf{y}$ is 0, the angle between the vectors is $\pi/2$ radians (perpendicular or orthogonal).

Calculating Dissimilarity between Vectors

Calculating Dissimilarity between Vectors

We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

Calculating Dissimilarity between Vectors

We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

1. calculating the distance dissimilarity as the norm of the difference vector $\|\mathbf{x} - \mathbf{y}\|_2$,

Calculating Dissimilarity between Vectors

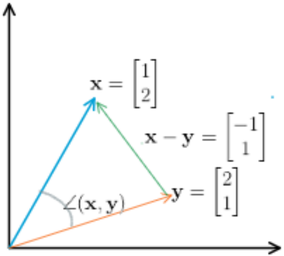
We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

1. calculating the distance dissimilarity as the norm of the difference vector $\|\mathbf{x} - \mathbf{y}\|_2$,
2. calculating the angular dissimilarity as $1 - \cos(\angle(\mathbf{x}, \mathbf{y}))$.

Calculating Dissimilarity between Vectors

We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

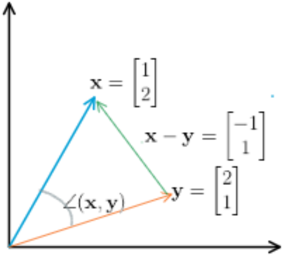
1. calculating the distance dissimilarity as the norm of the difference vector $\|\mathbf{x}-\mathbf{y}\|_2$,
2. calculating the angular dissimilarity as $1-\cos(\angle(\mathbf{x},\mathbf{y}))$.



Calculating Dissimilarity between Vectors

We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

1. calculating the distance dissimilarity as the norm of the difference vector $\|\mathbf{x} - \mathbf{y}\|_2$,
2. calculating the angular dissimilarity as $1 - \cos(\angle(\mathbf{x}, \mathbf{y}))$.

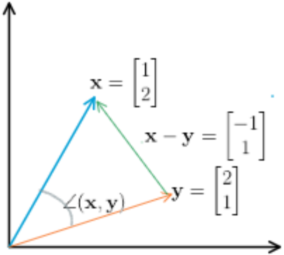


- Angular dissimilarity is the preferred choice when the vectors have a large number of components (a typical scenario in natural language processing).

Calculating Dissimilarity between Vectors

We can measure how dissimilar two vectors $\mathbf{x}$ and $\mathbf{y}$ are by

1. calculating the distance dissimilarity as the norm of the difference vector $\|\mathbf{x} - \mathbf{y}\|_2$,
2. calculating the angular dissimilarity as $1 - \cos(\angle(\mathbf{x}, \mathbf{y}))$.



- Angular dissimilarity is the preferred choice when the vectors have a large number of components (a typical scenario in natural language processing).
- When calculating distance between sample vectors (such as 2 patients), units for the features have to be considered.

Calculating Dissimilarity between Vectors–continued

Calculating Dissimilarity between Vectors–continued

Consider the following vectors whose elements are the house area (in thousands of sq. feet) and the no. of bedrooms for 3 houses:

Calculating Dissimilarity between Vectors–continued

Consider the following vectors whose elements are the house area (in thousands of sq. feet) and the no. of bedrooms for 3 houses:

$$\mathbf{x}^{(\mathbf{1})} = \begin{bmatrix} 1.6 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(\mathbf{2})} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(\mathbf{3})} = \begin{bmatrix} 1.6 \\ 4 \end{bmatrix}$$

Calculating Dissimilarity between Vectors–continued

Consider the following vectors whose elements are the house area (in thousands of sq. feet) and the no. of bedrooms for 3 houses:

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1.6 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(2)} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(3)} = \begin{bmatrix} 1.6 \\ 4 \end{bmatrix}$$

- $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 = 0.1$ is less than $\|\mathbf{x}^{(1)} - \mathbf{x}^{(3)}\|_2^2 = 2$ which indicates that house-1 similar to house-2.

Calculating Dissimilarity between Vectors–continued

Consider the following vectors whose elements are the house area (in thousands of sq. feet) and the no. of bedrooms for 3 houses:

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1.6 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(2)} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(3)} = \begin{bmatrix} 1.6 \\ 4 \end{bmatrix}$$

- $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 = 0.1$ is less than $\|\mathbf{x}^{(1)} - \mathbf{x}^{(3)}\|_2^2 = 2$ which indicates that house-1 similar to house-2.
- That is not the case if area is measured directly in square feet because then $\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 = 100$ is greater than $\|\mathbf{x}^{(1)} - \mathbf{x}^{(3)}\|_2^2 = 2$ which correctly indicates that house-1 similar to house-3.

Calculating Dissimilarity between Vectors–continued

Consider the following vectors whose elements are the house area (in thousands of sq. feet) and the no. of bedrooms for 3 houses:

$$\mathbf{x}^{(\mathbf{1})} = \begin{bmatrix} 1.6 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(\mathbf{2})} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, \quad \mathbf{x}^{(\mathbf{3})} = \begin{bmatrix} 1.6 \\ 4 \end{bmatrix}$$

- $\|\mathbf{x}^{(\mathbf{1})} - \mathbf{x}^{(\mathbf{2})}\|_2^2 = 0.1$ is less than $\|\mathbf{x}^{(\mathbf{1})} - \mathbf{x}^{(\mathbf{3})}\|_2^2 = 2$ which indicates that house-1 similar to house-2.
- That is not the case if area is measured directly in square feet because then $\|\mathbf{x}^{(\mathbf{1})} - \mathbf{x}^{(\mathbf{2})}\|_2^2 = 100$ is greater than $\|\mathbf{x}^{(\mathbf{1})} - \mathbf{x}^{(\mathbf{3})}\|_2^2 = 2$ which correctly indicates that house-1 similar to house-3.
- Standardizing the vectors (subtracting each feature from its mean and dividing by its standard deviation) prior to calculating distance dissimilarity will address this issue.

Matrix-Vector Product

Matrix-Vector Product

The dot product between two vectors can be easily extended to define the product of a matrix and vector:

Matrix-Vector Product

The dot product between two vectors can be easily extended to define the product of a matrix and vector:

$$\mathbf{A}\mathbf{x} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}}$$

Matrix-Vector Product

The dot product between two vectors can be easily extended to define the product of a matrix and vector:

$$\mathbf{Ax} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}} = \begin{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}}_{\text{row } 1} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \\ \underbrace{\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}}_{\text{row } 2} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \end{bmatrix}$$

Matrix-Vector Product

The dot product between two vectors can be easily extended to define the product of a matrix and vector:

$$\mathbf{Ax} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}} = \begin{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}}_{\text{row } 1} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \\ \underbrace{\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}}_{\text{row } 2} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 \times 4 + 2 \times 2 + 4 \times -2 \\ 2 \times 4 + 2 \times -1 + 3 \times -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Matrix-Vector Product

The dot product between two vectors can be easily extended to define the product of a matrix and vector:

$$\mathbf{A}\mathbf{x} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}} = \begin{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}}_{\text{row 1}} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \\ \underbrace{\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}}_{\text{row 2}} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 \times 4 + 2 \times 2 + 4 \times -2 \\ 2 \times 4 + 2 \times -1 + 3 \times -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

A matrix-vector product can be seen as a sequence of dot products between the rows of matrix $\mathbf{A}$ (seen as vectors) and the vector $\mathbf{x}$.

Matrix-Vector Product–continued

Matrix-Vector Product–continued

Surprisingly, the matrix-vector product can also be seen as a linear combination of the columns of matrix $\mathbf{A}$ using the elements of the vector $\mathbf{x}$ as multipliers:

Matrix-Vector Product–continued

Surprisingly, the matrix-vector product can also be seen as a linear combination of the columns of matrix $\mathbf{A}$ using the elements of the vector $\mathbf{x}$ as multipliers:

$$\mathbf{A}\mathbf{x} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}}$$

Matrix-Vector Product–continued

Surprisingly, the matrix-vector product can also be seen as a linear combination of the columns of matrix $\mathbf{A}$ using the elements of the vector $\mathbf{x}$ as multipliers:

$$\mathbf{A}\mathbf{x} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 2 \\ -1 \end{bmatrix} + -2\begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

Matrix-Vector Product–continued

Surprisingly, the matrix-vector product can also be seen as a linear combination of the columns of matrix $\mathbf{A}$ using the elements of the vector $\mathbf{x}$ as multipliers:

$$
\begin{aligned}
\mathbf{A}\mathbf{x} &= \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}} \underbrace{\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}}_{3\text{-vector}} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 2 \\ -1 \end{bmatrix} + -2\begin{bmatrix} 4 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 8 \end{bmatrix} + \begin{bmatrix} 4 \\ -2 \end{bmatrix} + \begin{bmatrix} -8 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\end{aligned}
$$

Matrix-Vector Product Examples

Matrix-Vector Product Examples

- Suppose the $m \times n$-matrix P gives the prices of n goods from m suppliers and q is an n-vector of quantities of the n goods;

Matrix-Vector Product Examples

- Suppose the $m \times n$-matrix $\mathbf{P}$ gives the prices of n goods from m suppliers and q is an n-vector of quantities of the n goods; then, $\mathbf{c} = \mathbf{P}\mathbf{q}$ is an m-vector that gives the total purchase cost from each of the m suppliers.

Matrix-Vector Product Examples

- Suppose the $m \times n$-matrix $\mathbf{P}$ gives the prices of n goods from m suppliers and q is an n-vector of quantities of the n goods; then, $\mathbf{c} = \mathbf{Pq}$ is an m-vector that gives the total purchase cost from each of the m suppliers.
- Suppose $\mathbf{X}$ is an $m \times n$-matrix representing n features of m patients and w is an n-vector of weights to be applied to the n features ;

Matrix-Vector Product Examples

- Suppose the $m \times n$-matrix $\mathbf{P}$ gives the prices of n goods from m suppliers and q is an n-vector of quantities of the n goods; then, $\mathbf{c} = \mathbf{P}\mathbf{q}$ is an m-vector that gives the total purchase cost from each of the m suppliers.
- Suppose $\mathbf{X}$ is an $m \times n$-matrix representing n features of m patients and w is an n-vector of weights to be applied to the n features ; then, $\mathbf{s} = \mathbf{X}\mathbf{w}$ is an m-vector that gives the weighted score of each of the m patients.

Matrix-Matrix Product

Matrix-Matrix Product

The product of two matrices is a natural and repeated extension of a matrix-vector product:

Matrix-Matrix Product

The product of two matrices is a natural and repeated extension of a matrix-vector product:

$$\mathbf{A} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}}, \quad \mathbf{B} = \underbrace{\begin{bmatrix} 4 & -1 \\ 2 & 0 \\ -2 & 3 \end{bmatrix}}_{3\times 2\text{-matrix}}$$

Matrix-Matrix Product

The product of two matrices is a natural and repeated extension of a matrix-vector product:

$$\mathbf{A} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}}, \; \mathbf{B} = \underbrace{\begin{bmatrix} 4 & -1 \\ 2 & 0 \\ -2 & 3 \end{bmatrix}}_{3\times 2\text{-matrix}}$$

$$\Rightarrow \mathbf{AB} = \begin{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \end{bmatrix}$$

Matrix-Matrix Product

The product of two matrices is a natural and repeated extension of a matrix-vector product:

$$\mathbf{A} = \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \end{bmatrix}}_{2\times 3\text{-matrix}}, \ \mathbf{B} = \underbrace{\begin{bmatrix} 4 & -1 \\ 2 & 0 \\ -2 & 3 \end{bmatrix}}_{3\times 2\text{-matrix}}$$

$$\Rightarrow \mathbf{AB} = \begin{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 11 \\ 0 & 7 \end{bmatrix}}_{2\times 2\text{-matrix}}.$$

Matrix-Matrix Product–continued

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB}$$

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a^{(1)}} \cdot \mathbf{b_1} & \mathbf{a^{(1)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(1)}} \cdot \mathbf{b_p} \\ \vdots & \dots & \dots & \vdots \\ \mathbf{a^{(m)}} \cdot \mathbf{b_1} & \mathbf{a^{(m)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(m)}} \cdot \mathbf{b_p} \end{bmatrix}$$

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a^{(1)}} \cdot \mathbf{b_1} & \mathbf{a^{(1)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(1)}} \cdot \mathbf{b_p} \\ \vdots & \dots & \dots & \vdots \\ \mathbf{a^{(m)}} \cdot \mathbf{b_1} & \mathbf{a^{(m)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(m)}} \cdot \mathbf{b_p} \end{bmatrix} = \begin{bmatrix} \mathbf{Ab_1} & \mathbf{Ab_2} & \dots & \mathbf{Ab_p} \end{bmatrix}.$$

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a^{(1)}} \cdot \mathbf{b_1} & \mathbf{a^{(1)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(1)}} \cdot \mathbf{b_p} \\ \vdots & \dots & \dots & \vdots \\ \mathbf{a^{(m)}} \cdot \mathbf{b_1} & \mathbf{a^{(m)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(m)}} \cdot \mathbf{b_p} \end{bmatrix} = \begin{bmatrix} \mathbf{Ab_1} & \mathbf{Ab_2} & \dots & \mathbf{Ab_p} \end{bmatrix}.$$

Suppose $\mathbf{X}$ is an $m \times n$-matrix representing n features of m patients

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a^{(1)}} \cdot \mathbf{b_1} & \mathbf{a^{(1)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(1)}} \cdot \mathbf{b_p} \\ \vdots & \dots & \dots & \vdots \\ \mathbf{a^{(m)}} \cdot \mathbf{b_1} & \mathbf{a^{(m)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(m)}} \cdot \mathbf{b_p} \end{bmatrix} = \begin{bmatrix} \mathbf{Ab_1} & \mathbf{Ab_2} & \dots & \mathbf{Ab_p} \end{bmatrix}.$$

Suppose $\mathbf{X}$ is an $m \times n$-matrix representing n features of m patients and $\mathbf{W}$ is an $n \times p$-matrix of weights, where each column of $\mathbf{W}$ corresponds to weights to be applied to the n features for a particular type of disease among p disease types;

Matrix-Matrix Product–continued

In general, if $\mathbf{A}$ is an $m \times n$-matrix and $\mathbf{B}$ is an $n \times p$-matrix, then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a^{(1)}} \cdot \mathbf{b_1} & \mathbf{a^{(1)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(1)}} \cdot \mathbf{b_p} \\ \vdots & \dots & \dots & \vdots \\ \mathbf{a^{(m)}} \cdot \mathbf{b_1} & \mathbf{a^{(m)}} \cdot \mathbf{b_2} & \dots & \mathbf{a^{(m)}} \cdot \mathbf{b_p} \end{bmatrix} = \begin{bmatrix} \mathbf{Ab_1} & \mathbf{Ab_2} & \dots & \mathbf{Ab_p} \end{bmatrix}.$$

Suppose $\mathbf{X}$ is an $m \times n$-matrix representing n features of m patients and $\mathbf{W}$ is an $n \times p$-matrix of weights, where each column of $\mathbf{W}$ corresponds to weights to be applied to the n features for a particular type of disease among p disease types; then, $\mathbf{S} = \mathbf{XW}$ is an $m \times p$-matrix whose columns give the weighted score of each of the m patients for each type of disease.

Hadamard Product

Hadamard Product

The Hadamard product (denoted as $\odot$) of two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes $(m \times n)$ is an elementwise product resulting in a matrix of the same size:

Hadamard Product

The Hadamard product (denoted as $\odot$) of two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes ($m \times n$) is an elementwise product resulting in a matrix of the same size:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \dots & a_{mn}b_{mn} \end{bmatrix}.$$

Hadamard Product

The Hadamard product (denoted as $\odot$) of two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes $(m \times n)$ is an elementwise product resulting in a matrix of the same size:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \dots & a_{mn}b_{mn} \end{bmatrix}.$$

The Hadamard product shows up in deep learning computations to simplify certain representations involving matrices and vectors.