CISC 371 Class 11

Linear Algebra – Kronecker Product and Hadamard Operators

Texts: [1] pp.242–244, 254–255

Main Concepts:
- *Vectorization and reshaping of a matrix*
- *Kronecker product of two matrices*
- *Kronecker-vectorization theorem*
- *Hadamard product is entrywise multiplication*

Sample Problem, Machine Learning: How can we concisely represent linear computations of neural networks?

In current usage, an *artificial neuron* is a mathematical model of a kind of biological cell that is found in the central nervous system of vertebrate animals. Of the many mathematical models that are in current applications, we will use an affine transformation of the inputs that are altered by a "sigmoid" activation function. Each entry of an input $\vec{x}_j \in \mathbb{R}^n$ is scaled by the corresponding entry of the weight vector $\hat{w} \in \mathbb{R}^n$, to which is added a bias scalar $b \in \mathbb{R}$; this scalar b is transferred by an activation function ϕ to produce an output value z. The computations will be described in further detail in Class 12.

For the classes on neural networks, we will depart slightly from our usual formalism. Our mathematical model of a neuron has "external" inputs, with the j^{th} data vector being $\vec{x}_j \in \mathbb{R}^n$, and "internal" variables such as a weight vector $\hat{w} \in \mathbb{R}^n$ plus a bias scalar $b \in \mathbb{R}$. The weights and bias are used to compute a hyper-plane scalar u_j as

$$u_j \stackrel{\text{def}}{=} \vec{x}_j \cdot \hat{w} + b \tag{11.1}$$

Most models of an artificial neuron incorporate a nonlinear *activation function*. We will write this as the function $\phi : \mathbb{R} \to \mathbb{R}$, so that the activation for the j^{th} data vector uses the hyper-plane scalar u_j of Equation 11.1 and is

$$\phi_j \stackrel{\text{def}}{=} \phi(u_j) \tag{11.2}$$

Our interest in neural networks is limited to using the in tasks for binary classification. Each data vector $\vec{x}_j$ will have a *label*, written as y_j, that specifies the data vector as being in exactly one of two sets. In much of machine learning, the sets are specified numerically as ± 1; however, in neural networks, it is common to use the specification of each label as

$$y_j \in \{0 \quad 1\} \tag{11.3}$$

11.1 Vectorization and Reshaping

A matrix that has m rows and n columns is a member of a vector space. This can be verified by establishing that such a matrix satisfies all of the requirements, or axioms, of a vector space. Concisely: for any two matrices $A \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{m \times n}$, and any real number $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$, the weighted sum is a matrix of the same size

$$[\alpha A + \beta C] \in \mathbb{R}^{m \times n} \tag{11.4}$$

If a matrix is in a vector space, then it may be convenient to think of transforming a matrix into a vector for conventional uses. One transformation is the process of *vectorization*. We can represent a matrix as a set of n columns, each column being a vector of size m, as

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \tag{11.5}$$

The columns of the matrix in Equation 11.5 can be "stacked" in a single vector that is of size (mn) so that

$$\text{vec}(A) \stackrel{\text{def}}{=} \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_n \end{bmatrix} \tag{11.6}$$

For example, we might have matrices A and C that are

$$A = \begin{bmatrix} 2 & 7 \\ 3 & 11 \\ 5 & 13 \end{bmatrix} \qquad C = \begin{bmatrix} 2 & 5 & 11 \\ 3 & 7 & 13 \end{bmatrix} \tag{11.7}$$

The matrices of Equation 11.7 have the same vectorization, which is

$$\text{vec}(A) = \text{vec}(C) = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \\ 11 \\ 13 \end{bmatrix} \tag{11.8}$$

For a vector that has a size with an appropriate integer factorization, the process of vectorization in Equation 11.6 can be reversed – but not uniquely. In `MATLAB` and other computer languages, the “reshape” operator has a vector and two sizes as arguments. The operator’s result is a matrix. An example is

$$\text{reshape}(\vec{a}, p, q) \Rightarrow A \in \mathbb{R}^{p \times q} \tag{11.9}$$

We will use the vectorization of Equation 11.6 sparingly and the reshaping of Equation 11.9 rarely.

11.2 The Kronecker Product $\otimes$

The *Kronecker product* of two matrices is also called the *direct product* or the *tensor product*. We will use operator to write expressions for multi-layer neural networks in a concise way.

The Kronecker product can be though of as “bringing” the second matrix into each entry of the first matrix. As a simple example, suppose that a matrix is a block diagonal matrix, with the same block repeated. In a matrix that has 2×2 such blocks, for a matrix $C \in \mathbb{R}^{p \times q}$, the block-diagonal matrix might look like

$$\begin{bmatrix} C & 0 \\ 0 & C \end{bmatrix} \tag{11.10}$$

The matrix in Equation 11.10 can be written as a Kronecker product of a 2×2 identity matrix and a matrix C, as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes C = \begin{bmatrix} 1\,C & 0 \\ 0 & 1\,C \end{bmatrix} \tag{11.11}$$

This concept that is exemplified in Equation 11.11, of bringing one matrix into entries of another, underlies the definition of the Kronecker product.

Definition: Kronecker product of two matrices

For any $A \in \mathbb{R}^{m \times n}$, and any $C \in \mathbb{R}^{p \times q}$, the *Kronecker product* $A \otimes C$ is defined as the matrix $M \in \mathbb{R}^{mp \times nq}$ that has partitions described as

$$m_{ij} \stackrel{\text{def}}{=} a_{ij} C \tag{11.12}$$

some of the many useful properties of the Kronecker product are described by Horn and Johnson [1]. These properties include the Kronecker product being, among other properties: commutative with scalars; associative; and distributive with addition. For matrices that have real numbers as entries, a transpose property is

$$[A \otimes C]^T = A^T \otimes C^T \tag{11.13}$$

A fundamental property of the Kronecker product is how it interacts with the ordinary matrix product.

Theorem: Mixed-product property of the Kronecker product

For any given $A \in \mathbb{R}^{m \times n}$, any given $M \in \mathbb{R}^{p \times q}$, any given $C \in \mathbb{R}^{n \times k}$, and any given $W \in \mathbb{R}^{q \times r}$,

$$[A \otimes M][C \otimes W] = [AC] \otimes [MW] \tag{11.14}$$

Proof: See Horn and Johnson [1], page 244.

Observation: We can apply the equation of Theorem 11.14 to solve simpler problems; for example, we can write $A[I \otimes W] = [A \otimes 1][I \otimes W]$ and proceed to find that $A[I \otimes W] = A \otimes W$.

11.3 The Kronecker-Vectorization Theorem

One important use of the Kronecker product is to solve a matrix extension of the ordinary linear equation $A\vec{w} = \vec{c}$, where A and $\vec{c}$ are given and the vector $\vec{w}$ is to be computed.

A matrix extension of the simple linear equation is: given A and C, find a matrix W that satisfies $AW = C$. This can be computed by solving a sequence of column problems, each of the form $A\vec{w}_j = \vec{c}_j$, but the Kronecker product provides a concise way of stating and equivalent computation.

Theorem: Kronecker vectorization of a matrix equation

For any given $A \in \mathbb{R}^{m \times n}$, any given $M \in \mathbb{R}^{p \times q}$, any given $C \in \mathbb{R}^{m \times q}$, and any unknown $W \in \mathbb{R}^{n \times p}$, the equation

$$AWM = C$$

is equivalent to a linear equation with the Kronecker product:

$$AWM = C \equiv [M^T \otimes A] \, \text{vec}(W) = \text{vec}(C) \tag{11.15}$$

Proof: See Horn and Johnson [1], page 254–255.

Observation: The vectorizations provide equivalent computations to a column-by-column method, but in a concise manner.

Corollary: mixed Kronecker matrix-vector product

For any given $A \in \mathbb{R}^{m \times n}$, any given $C \in \mathbb{R}^{n \times p}$, and any given $M \in \mathbb{R}^{p \times q}$,

$$[M \otimes A]\text{vec}(C) = \text{vec}(ACM^T) \tag{11.16}$$

Proof: This rewriting Equation 11.15

Observation: The vectorization of Equation 11.16 provides a way to concisely compute a matrix-vector term that involvec a Kronecker product as a simpler vectorization of a matrix term.

We can use the Kronecker product to write the linear terms of a uniform layer of artificial neurons in a neural network.

11.4 The Hadamard Product $\odot$ and Hadamard Operators

The Hadamard product is typically defined for a pair of matrices that each have m rows and n columns. For matrices $A \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{m \times n}$,

Definition: Hadamard product of two matrices

For any $A \in \mathbb{R}^{m \times n}$, and any $C \in \mathbb{R}^{m \times n}$, the *Hadamard product* $A \odot C$ is defined as the matrix $M \in \mathbb{R}^{m \times n}$ that has entries

$$m_{ij} \stackrel{\text{def}}{=} a_{ij} c_{ij} \tag{11.17}$$

The Hadamard product in Definition 11.17 is also called the *entrywise product* and, especially in the context of some programming languages, the *dot-times operator*.

The Hadamard product can be seen to apply to vectors that are represented as $\mathbb{R}^{m\times 1}$ matrices and to 1-forms that are represented as $\mathbb{R}^{1\times n}$ matrices. In some programming languages, such as `MATLAB`, the Hadamard product can be computed for a matrix and a compatible vector or 1-form. The matrix-vector and matrix-1-form operations can be defined as abbreviations for matrix operations.

Definition: Hadamard product of two matrices

For any $A \in \mathbb{R}^{m\times n}$ and any $\underline{u} \in \mathbb{R}^{1\times n}$ and any $\vec{v} \in \mathbb{R}^{m\times 1}$, the *Hadamard products* $A \odot \underline{u}$ and $A \odot \vec{v}$ are defined as

$$\begin{aligned} A \odot \underline{u} &\stackrel{\text{def}}{=} A \odot \left[\vec{1}\,\underline{u}\right] \\ A \odot \vec{v} &\stackrel{\text{def}}{=} A \odot \left[\vec{v}\,\underline{1}\right] \end{aligned} \tag{11.18}$$

11.4.1 Hadamard Properties

In expressions that contain only the Hadamard product, this operator is commutative, associative, and distributive. In expressions containing other operators, such as matrix multiplication, care may be needed to ensure that related factors are gathered correctly.

The Hadamard product of two vectors, $\vec{a} \odot \vec{c}$, is the same as the matrix-vector product that is the result of diagonalizing the first vector. Abbreviating $D_{\vec{a}}$ as the diagonal matrix that has a_i as entry (i, i), we can write

$$\begin{aligned} D_{\vec{a}} &\stackrel{\text{def}}{=} \text{diag}(\vec{a}) \\ \vec{a} \odot \vec{c} &= [D_{\vec{a}}]\vec{c} \end{aligned} \tag{11.19}$$

The Hadamard product is related to the Kronecker product by a mixed-product property.

Theorem: Mixed Hadamard-Kronecker products

For compatible matrices $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{m\times n}$ and $M \in \mathbb{R}^{p\times q}$ and $W \in \mathbb{R}^{p\times q}$, the Hadamard product of the Kronecker products is the Kronecker products of the Hadamard products, so

$$[A \otimes C] \odot [M \otimes W] = [A \odot C] \otimes [M \odot W] \tag{11.20}$$

Proof: by expansion.

11.4.2 Hadamard Operators

The Hadamard product of a matrix with itself, such as $A \odot A$, is a matrix in which each entry is the square of the corresponding entry of A. We can extend the Hadamard product to a Hadamard power, with the understanding that some operations may produce a complex number.

<u>Definition:</u> Hadamard power of a matrix

> For any $A \in \mathbb{R}^{m\times n}$ and any $s \in \mathbb{R}$, the *Hadamard power* of A to the power s, written as $A^{\circ s}$, is defined as the matrix $M \in \mathbb{C}^{m\times n}$ that has entries
>
> $$m_{ij} \stackrel{\text{def}}{=} (a_{ij})^s \tag{11.21}$$

The *Hadamard division* of two matrices is defined where the divisor matrix has non-zero entries.

<u>Definition:</u> Hadamard division of two matrices

> For any $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{m\times n}$ such that every $c_{ij} \neq 0$, the *Hadamard division* of A by C, written as $A \oslash C$, is defined as the matrix $M \in \mathbb{R}^{m\times n}$ that has entries
>
> $$m_{ij} \stackrel{\text{def}}{=} \frac{a_{ij}}{c_{ij}} \tag{11.22}$$

Correspondences of Hadamard operators and `MATLAB` expressions are provided in Table 11.1.

Table 11.1: Hadamard operators and corresponding `MATLAB` expressions, where all operands are compatible. Here, $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{m\times n}$ and $\underline{u} \in \mathbb{R}^{1\times n}$ and $\vec{v} \in \mathbb{R}^{m\times 1}$ and $s \in \mathbb{R}$.

Operation	**`MATLAB`**
$A \odot C$	`A.*C`
$A \odot \underline{u}$	`A.*u`
$A \odot \vec{v}$	`A.*v`
$A^{\circ s}$	`A.^s`
$A \oslash C$	`A./C`

References

[1] Horn RA, Johnson CR: Topics in Matrix Analysis. Cambridge University Press, 1991