Math for ML: review

Milos Hauskrecht
milos@pitt.edu
5329 Sennott Square, x4-8845

people.cs.pitt.edu/~milos/

ML and knowledge of other fields

ML solutions and algorithms rely on knowledge of many other disciplines:

- **Algebra**
- **Calculus**
- **Probability**
- **Statistics**
- **Control theory**
- **Decision theory**

Next: a review of the basics of algebra and calculus one typically needs for ML

Notation

Notation:

- **Scalar:** a

  Example: $a=3$

- **Vector:** **v** or $\vec{v}$

  Example:

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

Note: a vector is by default a column vector

- **Matrix:** M

  Example:

$$M = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix}$$

Terminology

- **Matrix:**

$$M = \begin{bmatrix} 2 & 6 \\ 1 & 8 \\ 9 & 10 \end{bmatrix}$$

3 x 2 matrix

Elements of the matrix:

$$M_{1,1} = 2, M_{1,2} = 6, \ldots, M_{3,2} = 10$$

$$U = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix}$$

3 x 3 matrix
(square matrix)

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

Vector is a special case of a matrix

Why vectors and matrices

Data: **Data instances are often represented using vectors, and datasets using matrices**.

Example: Weather information

Temperature	Pressure	Humidity	Cloud-cover	…
80	980	30	0 (clear)	
62	850	50	1 (partly cloudy)	
73	790	40	1 (partly cloudy)	
	….			

Data can be naturally represented as a matrix:

$$D = \begin{bmatrix} 80 & 980 & 30 & 0 \\ 62 & 850 & 50 & 1 \\ 73 & 790 & 40 & 1 \\ & \ldots & & \end{bmatrix}$$

Data instances in rows
Attributes in columns

Basic Operations

Matrix Transpose

- The transpose of a matrix is found by flipping the matrix over its main diagonal. The main diagonal is the diagonal that begins at the element located at the first row and first column of the matrix.

$$A = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix} \quad , \quad A^T = \begin{bmatrix} 2 & 1 & 8 \\ 6 & 8 & 10 \\ 4 & 9 & 3 \end{bmatrix}$$

Vector transpose:

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad , \quad \vec{v}^T = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix}$$

Basic Operations

Scalar/matrix operations: Addition

- Add the scalar to every element in the matrix. The sum of a matrix and a scalar is a matrix.

$$a = 3 \quad , \quad M = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix}$$

$$a + M = M + a = 3 + \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 9 & 7 \\ 4 & 11 & 12 \\ 11 & 13 & 6 \end{bmatrix}$$

Basic Operations

Scalar/matrix operations: Multiplication

- Multiply every element of the matrix by the scalar. The product of a matrix and a scalar is a matrix.

$$a = 3 \quad , \quad M = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix}$$

$$aM = Ma = 3 * \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 18 & 12 \\ 3 & 24 & 27 \\ 24 & 30 & 9 \end{bmatrix}$$

Basic Operations

Matrix/vector operations: Matrix Vector Addition

- Add the elements of the vector to each element in the corresponding row of the matrix. The sum of a vector and a matrix is a matrix.

$$\mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ -5 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\mathbf{v} + M = M + \mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ -5 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 10 & 9 & 8 \\ 2 & 3 & 4 \end{bmatrix}$$

Basic Operations

Matrix operations: Matrix-Matrix Addition

- Add the corresponding elements in the matrices together. <u>The matrices must be the same size.</u> The sum of two matrices is a matrix of the same size.

$$A = \begin{bmatrix} 3 & -4 & 5 \\ 10 & 9 & -8 \\ -2 & 3 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 3 & -4 & 5 \\ 10 & 9 & -8 \\ -2 & 3 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 8 \\ 16 & 14 & -4 \\ 5 & 11 & 14 \end{bmatrix}$$

Basic Operations

Matrix operations: Matrix-Matrix Multiplication

- The inner dimensions of the two matrices must be the same. The product matrix will have the same number of rows as the first matrix and the same number of columns as the second matrix.

Inner dimensions must agree | **Outer dimensions define the result**

3 x 2 | 2 x 3 | 3 x 3

$$A=\begin{bmatrix}3 & -4\\ 10 & 9\\ -2 & 3\end{bmatrix},\quad B=\begin{bmatrix}1 & 2 & 3\\ 6 & 5 & 4\end{bmatrix}\qquad C=AB=\begin{bmatrix}-21 & -14 & -7\\ 64 & 65 & 66\\ 16 & 11 & 6\end{bmatrix}$$

Basic Operations

Matrix operations: Matrix -Matrix Multiplication

- When performing matrix multiplication, take the sum of the products of the elements in the row of the first matrix and the column of the second matrix.

$$AB=\begin{bmatrix}3 & -4\\ 10 & 9\\ -2 & 3\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\ 6 & 5 & 4\end{bmatrix}=\begin{bmatrix}3+(-24) & 6+(-20) & 9+(-16)\\ 10+54 & 20+45 & 30+36\\ (-2)+18 & (-4)+15 & (-6)+12\end{bmatrix}$$

$$=\begin{bmatrix}-21 & -14 & -7\\ 64 & 65 & 66\\ 16 & 11 & 6\end{bmatrix}$$

Basic Operations

Matrix operations: Matrix-Matrix Multiplication

- When performing matrix multiplication, take the sum of the products of the elements in the row of the first matrix and the column of the second matrix.

$$AB = \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix} = \begin{bmatrix} 3+(-24) & 6+(-20) & 9+(-16) \\ 10+54 & 20+45 & 30+36 \\ (-2)+18 & (-4)+15 & (-6)+12 \end{bmatrix}$$

$$= \begin{bmatrix} -21 & -14 & -7 \\ 64 & 65 & 66 \\ 16 & 11 & 6 \end{bmatrix}$$

Basic Operations

Matrix operations: Matrix-Matrix Multiplication

- The product of A and B is not equal to the product of B and A.

$$AB = \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix} = \begin{bmatrix} 3+(-24) & 6+(-20) & 9+(-16) \\ 10+54 & 20+45 & 30+36 \\ (-2)+18 & (-4)+15 & (-6)+12 \end{bmatrix} = \begin{bmatrix} -21 & -14 & -7 \\ 64 & 65 & 66 \\ 16 & 11 & 6 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 3+20+(-6) & (-4)+18+9 \\ 18+50+(-8) & (-24)+45+12 \end{bmatrix} = \begin{bmatrix} 17 & 23 \\ 60 & 33 \end{bmatrix}$$

Basic Operations

Matrix/Vector operations:

Matrix-Vector Multiplication

- Multiplication of a matrix and a vector is similar to matrix-matrix multiplication. The inner dimensions of the matrix and the vector must match.

$$A = \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

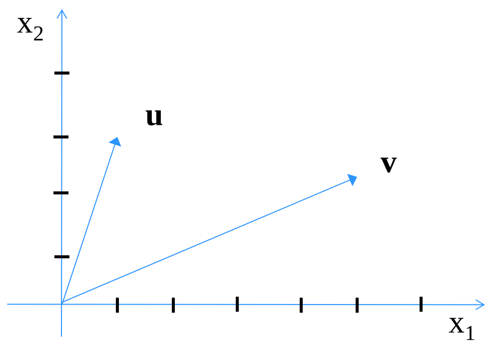
$$\mathbf{v}A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 17 & 23 \end{bmatrix} \qquad A\mathbf{u} = \begin{bmatrix} 3 & -4 \\ 10 & 9 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \end{bmatrix} = \begin{bmatrix} -19 \\ -4 \\ 14 \end{bmatrix}$$

Basic Operations

Matrix Vector Multiplication: geometric interpretation

- Assume a square matrix A and a vector u

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad \mathbf{v} = A\mathbf{u} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

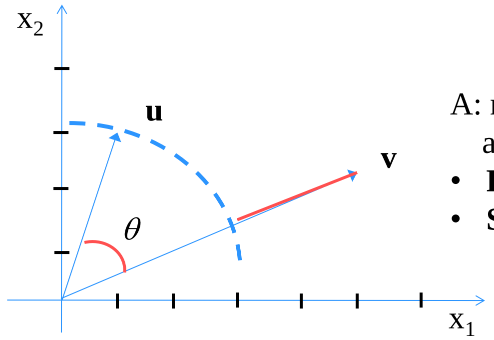


Basic Operations

Matrix Vector Multiplication: geometric interpretation

- Assume a square matrix A and a vector u

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad \mathbf{v} = A\mathbf{u} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$



A: represents
a combination of:

- **Rotation**
- **Stretching**

Basic Operations

Matrix/vector operations:

Vector-Vector Multiplication

- The product of two vectors of the same length is either a scalar or a matrix, depending on how the vectors are multiplied.

$$\mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 7 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 9 \\ 8 \\ 1 \end{bmatrix}$$

$$\mathbf{v}^T\mathbf{u} = \begin{bmatrix} 2 & -3 & 7 \end{bmatrix} \begin{bmatrix} 9 \\ 8 \\ 1 \end{bmatrix} = 18 + (-24) + 7 = 1$$

Inner (dot) product

$$\mathbf{v}\mathbf{u}^T = \begin{bmatrix} 2 \\ -3 \\ 7 \end{bmatrix} \begin{bmatrix} 9 & 8 & 1 \end{bmatrix} = \begin{bmatrix} 18 & 16 & 2 \\ -27 & -24 & -3 \\ 64 & 56 & 7 \end{bmatrix}$$

Outer product

Basic Operations

Matrix/Vector operations:

Matrix Inverse

- The product of a matrix and its inverse is the identity matrix

$$AA^{-1} = A^{-1}A = I$$

$$A = \begin{bmatrix} 3 & 0 & 2 \\ 2 & 0 & (-2) \\ 0 & 1 & 1 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 0.2 & 0.2 & 0 \\ (-0.2) & 0.3 & 1 \\ 0.2 & (-0.3) & 0 \end{bmatrix}$$

Note: The inverse of a matrix can be found by hand by augmenting the matrix with an identity matrix and using elementary row operations (additions or subtraction, multiplication by a constant, or swapping rows

Eigenvectors and Eigenvalues of the matrix

Decomposition allows us to see functional properties of a matrix

- The eigenvector of a square matrix M is a nonzero vector **v** such that, when multiplying the matrix M by the eigenvector, only the scale of the eigenvector changes.

$$M\mathbf{v} = \lambda\mathbf{v}$$

Eigenvalue (λ)   Eigenvector ($\mathbf{v}$)

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_1 = 3 \qquad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Matrix determinant

The determinant of a matrix maps matrices to real scalars

- The determinant is a measure of how much the space (vector) expands or contracts when multiplied by the matrix.

Examples:

$$A = \begin{bmatrix} 2 & 6 \\ 1 & 8 \end{bmatrix} \qquad \det(A) = 2*8 - 6*1 = 10$$

$$M = \begin{bmatrix} 2 & 6 & 4 \\ 1 & 8 & 9 \\ 8 & 10 & 3 \end{bmatrix}$$

$$\det(M) = 2(8*3 - 9*10) - 6(1*3 - 9*8) + 4(1*10 - 8*8)$$

Norms

A norm measures the size of a vector. It is a map from a vector to a non-negative scalar.

Properties of a norm:

- **1.** $f(\mathbf{x}) = 0 \quad \Rightarrow \quad \mathbf{x} = 0$
- **2.** $f(\mathbf{x}+\mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ **triangle inequality**
- **3.** $\forall a \in R \quad f(a\mathbf{x}) = |a| \, f(\mathbf{x})$

Examples of norms:

- **Euclidean (l_2 norm)**

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$$

Norms

Examples of norms:

- **Squared Euclidean (squared *l_2* norm*)***

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{d} x_i^2$$

- ***l_1* norm**

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{d} |x_i|$$

- **Max norm *(l infinity* norm)**

$$\|\mathbf{x}\|_\infty = \max_i |x_i|$$

Functions

Functions of one variable:

$$f(x) = x^2$$

$$f(x) = \log x$$

Function of many variables

$$f(x_1, x_2) = x_1^2 + x_2$$

$$f(x_1, x_2, x_3) = x_1^2 + x_2 + 2x_3 + 3$$

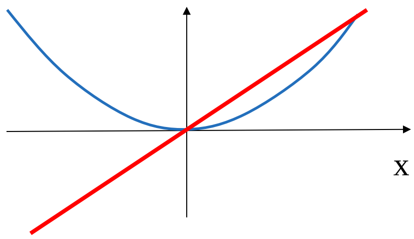
Function derivatives

Function derivatives are useful to analyze functions and their behaviors:

- **First derivatives: increasing, decreasing trends and extremes**

$$f(x) = x^2$$

$$f(x)' = \frac{dx^2}{\partial x} = 2x$$



- $2x$ **at** x=2   4 (increasing)
- $2x$ **at** x=-3 -6 (decreasing)
- $2x$ **at** x= 0   0 (an extreme - minimum)
- **Solving for** $f(x)' = 0$ **helps us to find the function extremes**

Function derivatives

The same applies for multivariate functions

- **First derivatives: increasing, decreasing trends and extremes**

$$\mathbf{x} = [x_1, x_2]^T$$

$$f(\mathbf{x}) = x_1^{\ 2} + x_2$$

- **Gradient – a vector of partial derivatives (Jacobian)**

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix}$$

- **Solving for** $\nabla_{\mathbf{x}} f(\mathbf{x}) = \overline{0}$ **helps us to find the function extremes**

Function derivatives

The same applies for multivariate functions

- **First derivatives: increasing, decreasing trends and extremes**

$$\mathbf{x} = [x_1, x_2]^T$$

$$f(\mathbf{x}) = x_1^{\ 2} + x_2$$

- **Hessian – a matrix of second partial derivatives**

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial f(\mathbf{x})}{\partial x_1^{\ 2}} & \dfrac{\partial f(\mathbf{x})}{\partial x_1 \partial x_2} \\ \dfrac{\partial f(\mathbf{x})}{\partial x_2 \partial x_1} & \dfrac{\partial f(\mathbf{x})}{\partial x_2^{\ 2}} \end{bmatrix}$$

- **Describes the local curvature of the function**