Linear Algebra Review
(Adapted from Punit Shah's slides)

Introduction to Machine Learning (CSC 311)
Spring 2020

University of Toronto

Basics

- A scalar is a number.

- A vector is a 1-D array of numbers. The set of vectors of length n with real elements is denoted by $\mathbb{R}^n$.
  - Vectos can be multiplied by a scalar.
  - Vector can be added together if dimensions match.

- A matrix is a 2-D array of numbers. The set of $m \times n$ matrices with real elements is denoted by $\mathbb{R}^{m \times n}$.
  - Matrices can be added together or multiplied by a scalar.
  - We can multiply Matrices to a vector if dimensions match.

- In the rest we denote scalars with lowercase letters like a, vectors with bold lowercase $\mathbf{v}$, and matrices with bold uppercase $\mathbf{A}$.

Norms

- Norms measure how "large" a vector is. They can be defined for matrices too.

- The ℓ_p-norm for a vector $\mathbf{x}$:

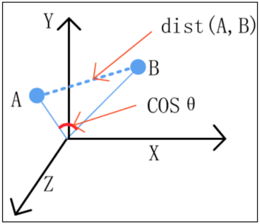
$$\|\mathbf{x}\|_p = \left[\sum_i |x_i|^p\right]^{\frac{1}{p}}.$$

  - The ℓ_2-norm is known as the Euclidean norm.
  - The ℓ_1-norm is known as the Manhattan norm, i.e., $\|\mathbf{x}\|_1 = \sum_i |x_i|$.
  - The ℓ_∞ is the max (or supremum) norm, i.e., $\|\mathbf{x}\|_\infty = \max_i |x_i|$.

Dot Product

- Dot product is defined as $\mathbf{v} \cdot \mathbf{u} = \mathbf{v}^\top \mathbf{u} = \sum_i u_i v_i$.
- The ℓ_2 norm can be written in terms of dot product: $\|\mathbf{u}\|_2 = \sqrt{\mathbf{u}.\mathbf{u}}$.
- Dot product of two vectors can be written in terms of their ℓ_2 norms and the angle θ between them:

$$\mathbf{a}^\top \mathbf{b} \; = \|\mathbf{a}\|_2 \, \|\mathbf{b}\|_2 \cos(\theta).$$



Cosine Similarity

- Cosine between two vectors is a measure of their similarity:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \, \|\mathbf{b}\|}.$$

- **Orthogonal Vectors**: Two vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal to each other if $\mathbf{a} \cdot \mathbf{b} = 0$.

Vector Projection

- Given two vectors $\mathbf{a}$ and $\mathbf{b}$, let $\hat{\mathbf{b}} = \frac{\mathbf{b}}{\|\mathbf{b}\|}$ be the unit vector in the direction of $\mathbf{b}$.
- Then $\mathbf{a}_1 = a_1 \cdot \hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{a}$ onto a straight line parallel to $\mathbf{b}$, where

$$a_1 = \|\mathbf{a}\| \cos(\theta) = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{\|\mathbf{b}\|}$$

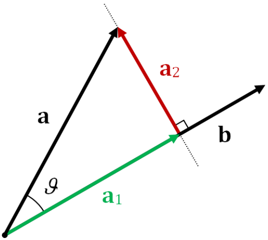


Image taken from wikipedia.

Trace

- Trace is the sum of all the diagonal elements of a matrix, i.e.,

$$\mathrm{Tr}(\mathbf{A}) = \sum_i A_{i,i}.$$

- Cyclic property:

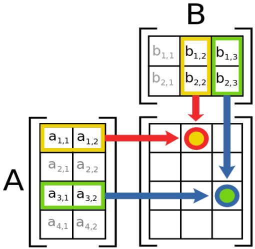
$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA}).$$

Multiplication

- Matrix-vector multiplication is a linear transformation. In other words,

$$\mathbf{M}(v_1 + av_2) = \mathbf{M}v_1 + a\mathbf{M}v_2 \implies (\mathbf{M}v)_i = \sum_j M_{i,j}v_j.$$

- Matrix-matrix multiplication is the composition of linear transformations, i.e.,
$(\mathbf{AB})v = \mathbf{A}(\mathbf{B}v) \implies (\mathbf{AB})_{i,j} = \sum_k A_{i,k}B_{k,j}.$



Invertibality

- $\mathbf{I}$ denotes the identity matrix which is a square matrix of zeros with ones along the diagonal. It has the property $\mathbf{IA} = \mathbf{A}$ ($\mathbf{BI} = \mathbf{B}$) and $\mathbf{Iv} = \mathbf{v}$

- A square matrix $\mathbf{A}$ is invertible if $\mathbf{A}^{-1}$ exists such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

- Not all non-zero matrices are invertible, e.g., the following matrix is not invertible:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Transposition

- Transposition is an operation on matrices (and vectors) that interchange rows with columns. $(\mathbf{A}^\top)_{i,j} = \mathbf{A}_{j,i}$.
- $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.
- $\mathbf{A}$ is called symmetric when $\mathbf{A} = \mathbf{A}^\top$.
- $\mathbf{A}$ is called orthogonal when $\mathbf{AA}^\top = \mathbf{A}^\top \mathbf{A} = \mathbf{I}$ or $\mathbf{A}^{-1} = \mathbf{A}^\top$.

Diagonal Matrix

- A diagonal matrix has all entries equal to zero except the diagonal entries which might or might not be zero, e.g. identity matrix.
- A square diagonal matrix with diagonal enteries given by entries of vector $\mathbf{v}$ is denoted by $\text{diag}(\mathbf{v})$.
- Multiplying vector $\mathbf{x}$ by a diagonal matrix is efficient:

$$\text{diag}(\mathbf{v})\mathbf{x} \;=\; \mathbf{v} \odot \mathbf{x},$$

  where $\odot$ is the entrywise product.

- Inverting a square diagonal matrix is efficient

$$\text{diag}(\mathbf{v})^{-1} \;=\; \text{diag}\Big([\frac{1}{v_1}, \ldots, \frac{1}{v_n}]^\top\Big).$$

Determinant

- Determinant of a square matrix is a mapping to scalars.

$$\det(\mathbf{A}) \quad \text{or} \quad |\mathbf{A}|$$

- Measures how much multiplication by the matrix expands or contracts the space.
- Determinant of product is the product of determinants:

$$\det(\mathbf{AB}) \; = \; \det(\mathbf{A})\det(\mathbf{B})$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \; = \; ad - bc$$

List of Equivalencies

Assuming that $\mathbf{A}$ is a square matrix, the following statements are equivalent

- $\mathbf{Ax} = \mathbf{b}$ has a **unique** solution (for every b with correct dimension).
- $\mathbf{Ax} = \mathbf{0}$ has a unique, trivial solution: $\mathbf{x} = \mathbf{0}$.
- Columns of $\mathbf{A}$ are linearly independent.
- $\mathbf{A}$ is invertible, i.e. $\mathbf{A}^{-1}$ exists.
- $\det(\mathbf{A}) \neq 0$

Zero Determinant

If $\det(\mathbf{A}) = 0$, then:

- $\mathbf{A}$ is linearly dependent.
- $\mathbf{Ax} = \mathbf{b}$ has infinitely many solutions or no solution. These cases correspond to when b is in the span of columns of $\mathbf{A}$ or out of it.
- $\mathbf{Ax} = \mathbf{0}$ has a non-zero solution. (since every scalar multiple of one solution is a solution and there is a non-zero solution we get infinitely many solutions.)