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 $\mathbb{R}^n$.
  - Vectors can be multiplied by a scalar.
  - Vectors 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 measure how “large” a vector is. They can be defined for matrices too.

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

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

- The $\ell_2$-norm is known as the Euclidean norm.
- The $\ell_1$-norm is known as the Manhattan norm, i.e., $\|\mathbf{x}\|_1 = \sum_i |x_i|$.
- The $\ell_\infty$ 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 $\ell_2$ norm can be written in terms of dot product: $\|\mathbf{u}\|_2 = \sqrt{\mathbf{u} \cdot \mathbf{u}}$.

- Dot product of two vectors can be written in terms of their $\ell_2$ norms and the angle $\theta$ 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:
  \[ \cos(\theta) = \frac{a \cdot b}{\|a\| \|b\|}. \]

- **Orthogonal Vectors:** Two vectors \(a\) and \(b\) are orthogonal to each other if \(a \cdot b = 0\).
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 $\mathbf{a}_1 = \mathbf{a}_1 \cdot \hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{a}$ onto a straight line parallel to $\mathbf{b}$, where

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

- Trace is the sum of all the diagonal elements of a matrix, i.e.,
  \[ \text{Tr}(A) = \sum_i A_{i,i}. \]

- Cyclic property:
  \[ \text{Tr}(ABC) = \text{Tr}(CAB) = \text{Tr}(BCA). \]
Multiplication

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

$$M(v_1 + a v_2) = Mv_1 + aMv_2 \implies (Mv)_i = \sum_j M_{i,j}v_j.$$

- Matrix-matrix multiplication is the composition of linear transformations, i.e.,

$$(AB)v = A(Bv) \implies (AB)_{i,j} = \sum_k A_{i,k}B_{k,j}.$$
Invertibility

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

- A square matrix \( A \) is invertible if \( A^{-1} \) exists such that \( A^{-1}A = AA^{-1} = I \).

- 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. \((A^\top)_{i,j} = A_{j,i}\).

- \((AB)^\top = B^\top A^\top\).

- \(A\) is called symmetric when \(A = A^\top\).

- \(A\) is called orthogonal when \(AA^\top = A^\top A = I\) or \(A^{-1} = A^\top\).
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 entries 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

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

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

\[
\det(A) \quad \text{or} \quad |A|
\]

- Measures how much multiplication by the matrix expands or contracts the space.

- Determinant of product is the product of determinants:

\[
\det(AB) = \det(A)\det(B)
\]

\[
\begin{vmatrix}
 a & b \\
 c & d
\end{vmatrix} = ad - bc
\]
List of Equivalencies

Assuming that $A$ is a square matrix, the following statements are equivalent

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

- **A** is linearly dependent.

- \( Ax = b \) has infinitely many solutions or no solution. These cases correspond to when \( b \) is in the span of columns of \( A \) or out of it.

- \( Ax = 0 \) 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.)