Mathematics for Machine Learning

1. System of Equation

Linear algebra provides a way of compactly representing and operating on sets of linear equations. For example, consider the following system of equations:

$$\begin{array}{rcrcl} 4x_1 & - & 5x_2 & = & -13 \\ -2x_1 & + & 3x_2 & = & 9. \end{array}$$

In matrix notation, we can write the system more compactly as Ax = b

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}.$$

- By A ∈ R m×n we denote a matrix with m rows and n columns, where the entries of A are real numbers.

- By x ∈ R n , we denote a vector with n entries. By convention, an n-dimensional vector is often thought of as a matrix with n rows and 1 column, known as a column vector. If we want to explicitly represent a row vector — a matrix with 1 row and n columns.

2. Matrix Multiplication

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

- Note that in order for the matrix product to exist, the number of columns in A must equal the number of rows in B.

3. Vector-Vector Products

Given two vectors x, y ∈ R, the inner product or dot product of the vectors, is a real number given by

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$

Given vectors x ∈ R (m), y ∈ R (n), xyT ∈ R m×n is called the outer product of the vectors.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

4. Matrix-Vector Product

- To define multiplication between a matrix M and a vector V, we need to view the vector as a column matrix.
- We define the matrix-vector product only for the case when the number of columns in M equals the number of rows in V.

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

5. Matrix-Matrix Product

- Matrix multiplication is associative: (AB)C = A(BC).
- Matrix multiplication is distributive: A(B + C) = AB + AC.
- Matrix multiplication is, in general, not commutative; that is, it can be the case that AB != BA.

6. The Identity Matrix and Diagonal Matrices

- The identity matrix, denoted is a square matrix with ones on the diagonal and zeros everywhere else.

  $$AI = A = IA$$

  Note that in some sense, the notation for the identity matrix is ambiguous, since it does not specify the dimension of I.
  Generally, the dimensions of I are inferred from context so as to make matrix multiplication possible.
  For example, in the equation above, the I in AI = A is an (n×n) matrix, whereas the I in A = IA is an (m×m) matrix.

- A diagonal matrix is a matrix where all non-diagonal elements are 0

7. Transpose

- (AT) T = A
- (AB) T = BTAT
- (A + B) T = AT + BT

8. Symmetric Matrices

- A square matrix is symmetric if A = AT

- It is anti-symmetric if A = −AT

- The matrix A + AT is symmetric and the matrix A−AT is anti-symmetric

- From this it follows that any square matrix can be represented as a sum of a symmetric matrix and an antisymmetric matrix.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

9. The Trace

The trace of a square matrix , denoted tr(A) (or just trA if the parentheses are obviously implied), is the sum of diagonal elements in the matrix.

- trA = trAT
- tr(A + B) = trA + trB
- tr(tA) = t trA
- For A, B such that AB is square, trAB = trBA
- For A, B, C such that ABC is square, trABC = trBCA = trCAB, and so on for the product of more matrices

10. Norms

- A norm of a vector ||x|| is informally a measure of the “length” of the vector.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

11. Linear Independence and Rank

- A set of vectors {x1, x2, . . . xn}⊂R m is said to be (linearly) independent if no vector can be represented as a linear combination of the remaining vectors.

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

*for scalar values of alpha.

- Conversely, if one vector belonging to the set can be represented as a linear combination of the remaining vectors, then the vectors are said to be (linearly) dependent.

- The **column rank** of a matrix A ∈ R (m×n) is the size of the largest subset of columns of A that constitute a linearly independent set.

- **Row rank** is the largest number of rows of A that constitute a linearly independent set

- It turns out that the column rank of A is equal to the row rank of A, and so both quantities are referred to collectively as the rank of A

*Please refer to the properties of rank of matrices.

12. The Inverse of a Square Matrix

- The inverse of a square matrix $A \in R$ (n×n) is denoted A^{-1} , and is the unique matrix such that:

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

- Non-square matrices, do not have inverses by definition. However, for some square matrices A, it may still be the case that 11 A^{-1} may not exist.

- In particular, we say that A is invertible or non-singular if $A-1$ exists.

- We say non-invertible or singular otherwise.

- In order for a square matrix A to have an inverse A^{-1} , then A must be full rank.

*Please refer to the properties of inverse

13. Eigenvalues and Eigenvectors

- Given a square matrix $A \in R(n \times n)$, we say that $\lambda \in C$ is an eigenvalue of A and $x \in C(n)$ is the corresponding eigenvector if

$$Ax = \lambda x, \quad x \neq 0.$$

- This definition means that multiplyingAby the vector x results in a new vector that points in the same direction as x, but scaled by a factor λ.

- For any eigenvector $x \in C(n)$, and scalar $t \in C$,
  $A(cx) = cAx = c\lambda x = \lambda(cx)$, so cx is also an eigenvector

- We can rewrite the equation above to state that (λ, x) is an eigenvalue-eigenvector pair of A if,

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

- But $(\lambda I - A)x = 0$ has a non-zero solution to x if and only if $(\lambda I - A)$ has a non-empty nullspace, which is only the case if $(\lambda I - A)$ is singular, i.e.

$$|(\lambda I - A)| = 0.$$