Linear Algebra – 1
Notes based on the book
Mathematics for Machine Learning

Recall the checkers game, where we came up with a target function of the type

$$w_0 + w_1 \cdot bp(b) + w_2 \cdot rp(b) + w_3 \cdot bk(b) + w_4 \cdot rk(b) + w_5 \cdot bt(b) + w_6 \cdot rt(b)$$

- $bp(b)$: number of black pieces on board b
- $rp(b)$: number of red pieces on b
- $bk(b)$: number of black kings on b
- $rk(b)$: number of red kings on b
- $bt(b)$: number of red pieces threatened by black (i.e., which can be taken on black's next turn)
- $rt(b)$: number of black pieces threatened by red

We can generalize the form of the above function by using a generic notation x for the game board and $x_1, \ldots x_6$ as the various features we use to describe the board. Therefore, one would obtain:

$$w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

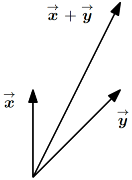
Notice now the similarity in writing the coefficients (weights) w_i and the corresponding features x_i, $i = 1, \ldots 6$. This is a very convenient way of writing this function. Moreover, notice that one can write this function in a more compact way as $w_0 + \sum_{i=1}^{6} w_ix_i$, and even more compact if we use the following "trick": extend the notation x_i, by defining $x_0 = 1$. Then

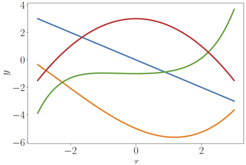
$$w_0 + \sum_{i=1}^{6} w_ix_i = w_0 \cdot 1 + \sum_{i=1}^{6} w_ix_i = w_0 \cdot x_0 + \sum_{i=1}^{6} w_ix_i = \sum_{i=0}^{6} w_ix_i$$

We will say that w and x are *vectors* of dimension 7: with components $w_0, \ldots w_6$ and $x_0 = 1, \ldots x_6$, respectively. The notions of vectors and matrices are very important in Machine Learning.

VECTORS – 1

In general, vectors are mathematical objects that can be added together and multiplied by scalars to produce another object of the same kind.





Geometric Vectors (left)
polynomials as Vectors (right)

In both cases addition and multiplication with a scalar are possible.

VECTORS – 2

In general any collection of like values can be represented as a vector:

- Audio signals, image pixel information, etc.
- Elements of $\mathbb{R}^n$, the n dimensional Euclidean space are vectors too, more abstract than, say polynomials.

If $n = 3$, then

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$$

In a computer, vectors are usually implemented as the data structure *array*.

VECTOR SPACES – 3

- **Question:** What is the set of vectors that can result by starting with a small set of vectors, and adding them to each other and scaling them?
- **Answer:** This results in a vector space and we say that a vector space is closed with respect to addition and multiplication with a scalar.

Find the optimal production plan, given

1. $N_1, \ldots N_n$ products are produced
2. $R_1, \ldots, R_m$ resources are required
3. To produce N_j, a_{ij} units of resource R_i are needed; $i = 1, \ldots, m$, $j = 1, \ldots, n$
4. Optimal Plan $= x_j$ of product N_j should be produced if a total of b_i units of resource R_i are available and (ideally) no resources are left over.

To solve this we need to solve the following *system of equations*

$$\begin{aligned} &a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ &\ldots \\ &a_{n1}x_1 + \cdots + a_{mn}x_n = b_m \end{aligned}$$

System of Linear Equations (no solution)

$$
\begin{array}{ccccccccc}
x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\
x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\
2x_1 & & & + & 3x_3 & = & 1 & (3)
\end{array}
$$

(1) + (2) obtains $2x_1 + 3x_3 = 5$ which contradicts (3).
Therefore, there is *no solution*

System of Linear Equations (unique solution)

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ & & x_2 & + & x_3 & = & 2 & (3) \end{array}$$

(1) & (3) implies that $x_1 = 1$;
(1)+(2) obtains $2x_1 + 3x_3 = 5$, that is, $2 + 3x_3 = 5$, that is $3x_3 = 3$ from which it follows that $x_3 = 1$.
Finally, from (3) we obtain $x_2 = 1$.
Hence the unique solution is $(1, 1, 1)$.

System of Linear Equations (infinitely many solutions)

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ & & x_2 & + & x_3 & = & 5 & (3) \end{array}$$

Since (1) + (2) = (3) equation (3) can be omitted.

From (1)+(2) obtains $2x_1 = 5 - 3x_3$, therefore $x_1 = \frac{1}{2}(5 - 3x_3)$.

From (1) - (2) obtains $2x_2 - x_3 = 1$, therefore, $x_2 = \frac{1}{2}(1 + x_3)$.

We define $x_3 = a \in \mathbb{R}$ and we notice that any triplet of the form

$$\left(\frac{1}{2}(5 - 3a), \frac{1}{2}(1 + a), a\right), \quad a \in \mathbb{R}$$

is a solution.

We conclude that in general, for a real-valued system of linear equations we obtain either (1) no solution, (2) exactly one solution, or (3) infinitely many solutions.

A well known, important Machine Learning procedure - Linear Regression solves a version of the examples we have seen above when we cannot solve the system of linear equations.

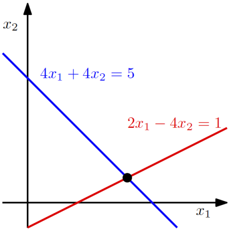
Geometric interpretation

... of solving a system of (linear) equations

Consider the system with two unknowns x_1 and x_2

$$\begin{array}{rcrcl} 4x_1 & + & 4x_2 & = & 5 \\ 2x_1 & - & 4x_2 & = & 1 \end{array}$$

Each of these equations represent a line in $\mathbb{R}^2$. The unique solution is $(x_1, x_2) = \left(1, \frac{1}{4}\right)$, which is in fact the point at which the two lines intersect.



Introduce compact notation: Collect coefficients a_{ij} into vectors; collect vectors into *matrices*. Then we can write:

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

or,

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

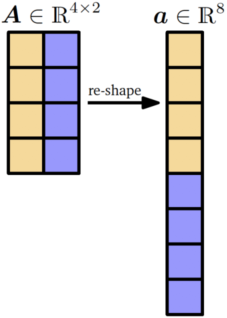
Notice that the number of columns in the matrix is equal to the number of row in the vector x.

We define a matrix as follows: with $m, n \in \mathbb{N}$ (the positive integers), the real valued (m, n) matrix $\mathbf{A}$ is an $m - tuple$ of elements a_{ij}, with $i = 1, \ldots, m, j = 1, \ldots, n$ which is ordered in a rectangular scheme consisting of m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, a_{ij} \in \mathbb{R}$$

If $m = 1$, the $(1, n)$ matrices are called *rows/ row vectors*;
If $n = 1$, the $(m, 1)$ matrices are called *columns/column vectors*.

We denote by $\mathbb{R}^{m \times n}$ the collection of all (m, n) matrices. Matrices can be stacked (reshaped) so that if $\mathbf{A} \in \mathbb{R}^{m \times n}$, then $\mathbf{A} \in \mathbb{R}^{mn}$. We usually read $\mathbf{A} \in \mathbb{R}^{m \times n}$ as '$\mathbf{A}$ is an *m by n matrix*'.



Recall listing the elements of a two-dimensional array in *column major order*.

Matrix addition & multiplication

- $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$: the sum, $\mathbf{C} = \mathbf{A} + \mathbf{B} \in \mathbb{R}^{m \times n}$ with elements $c_{ij} = a_{ij} + b_{ij}$.
- $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times k}$: the product, $\mathbf{C} = \mathbf{AB} \in \mathbf{R}^{m \times k}$, with elements $c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}$, with $i = 1, \ldots m, j = 1, \ldots k$.
- The product $\mathbf{BA}$ is not defined if $m \neq n$.
- Matrix multiplication is NOT DEFINED as element-wise multiplication. That is $c_{ij} \neq a_{ij} b_{ij}$ (this type of operation is called the *Hadamard Product* which has found many applications in statistics.

Exercise

Given $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$ and $\mathbf{B} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}$,

compute

- The product $\mathbf{AB} \in \mathbb{R}^{2\times 2}$
- The product $\mathbf{BA} \in \mathbb{R}^{3\times 3}$

Since $\mathbf{AB} \neq \mathbf{BA}$ we say that the matrix product is NOT COMMUTATIVE.

The Identity Matrix

$\mathbf{I}_n$ diagonal $n \times n$ matrix with diagonal elements 1 and 0*s* elsewhere:

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Properties of Matrices

- *Associativity*: $\forall \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}, \mathbf{C} \in \mathbb{R}^{p\times q}$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

- *Distributivity:* $\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{n\times p}$,

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AB} + \mathbf{BC}$$

$$\mathbf{A}(\mathbf{C} + \mathbf{D}) = \mathbf{AC} + \mathbf{AD}$$

- Multiplication with the identity matrix:

$$\forall \mathbf{A} \in \mathbb{R}^{m\times n} : \mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$$

Note that $\mathbf{I}_m \neq \mathbf{I}_n$, for $m \neq n$.

Inverse Matrix

$\mathbf{A} \in \mathbb{R}^{n \times n}$. The matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ with the property that

$$\mathbf{AB} = \mathbf{I}_n$$

is called the *inverse* of matrix $\mathbf{A}$. Note the similarity of the operations with real numbers, where the number 1 is what $\mathbf{I}$ is for matrices, and for a real value $a \neq 0$, its inverse, $\frac{1}{a}$ has the property that $a\left(\frac{1}{a}\right) = 1$.

Calculating the inverse of a matrix occurs very often in ML algorithms.

We will see that the inverse of a matrix does not always exists, in which case we often compute something called a pseudoinverse.

The inverse matrix of a 2×2 matrix

Let us consider the matrix

$$\mathbf{A} \in \mathbb{R}^{2\times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

We want to find the matrix $\mathbf{X} \in \mathbb{R}^{2\times 2}$ such that

$$\mathbf{AX} = \mathbf{XA} = \mathbf{I}_2$$

Assume $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$

Then

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22} \\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22} \end{bmatrix}$$

Set
$$\begin{bmatrix} a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22} \\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
that is, we obtain the linear system of 4 equations with 4 unknowns:
$$\begin{cases} a_{11}x_{11} + a_{12}x_{21} = 1 \\ a_{11}x_{12} + a_{12}x_{22} = 0 \\ a_{21}x_{11} + a_{22}x_{21} = 0 \\ a_{21}x_{12} + a_{22}x_{22} = 1 \end{cases}$$
The solution of this system is:
$$x_{ij} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}(-1)^{i+j}a_{ij} \iff D = a_{11}a_{22} - a_{12}a_{21} \neq 0,$$
where D is the determinant of the matrix $\mathbf{A}$.

Important Note: The determinant can be used to check whether the matrix is invertible (i.e., it has an inverse) or not. We will return to algorithms for solving such systems.

Transpose of a matrix

$\mathbf{A} \in \mathbb{R}^{n\times n}$ with elements a_{ij}. The matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ with elements $b_{ij} = a_{ji}$ is called transpose of $\mathbf{A}$ denoted $\mathbf{A}^\top$.

The transpose of a matrix is obtained by writing the rows of the transpose from the columns of the matrix.

Properties connecting the inverse and transpose matrices are as follows:

1. $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$
2. $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
3. $(\mathbf{A}+\mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1}$
4. $(\mathbf{A}^\top)^\top = \mathbf{A}$
5. $(\mathbf{A}+\mathbf{B})^\top \neq \mathbf{A}^\top + \mathbf{B}^\top$
6. $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top + \mathbf{A}^\top$

Symmetric Matrix

A square matrix, that is, $\mathbf{A} \in \mathbb{R}^{n\times n}$, is said to be symmetric if $\mathbf{A} = \mathbf{A}^\top$. For example, the matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \text{ is symmetric since its transpose is } \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

If $\mathbf{A}$ is invertible, then $\mathbf{A}^\top$ is also invertible and $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$ and it is denoted by $\mathbf{A}^{-\top}$. The sum of two symmetric matrices is symmetric:

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} + \begin{bmatrix} c & d \\ d & c \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ b+d & a+c \end{bmatrix}$$

However, this is <u>not necessarily true</u> for the multiplication of two symmetric matrices:

$$\begin{bmatrix} a & b \\ b & b \end{bmatrix} \times \begin{bmatrix} c & c \\ c & c \end{bmatrix} = \begin{bmatrix} ac+bc & ac+bc \\ 2bc & 2bc \end{bmatrix}$$

which is not equal to its transpose

$$\begin{bmatrix} ac+bc & 2bc \\ ac+bc & 2bc \end{bmatrix}$$

Multiplication with a scalar/properties

If $\lambda \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{m \times n}$, with elements $a_{ij}, i = 1, \ldots, m, j = 1, \ldots n$, then $\mathbf{B} = \lambda \mathbf{A} \in \mathbb{R}^{m \times n}$,has elements $b_{ij} = \lambda a_{ij}$. Let $\lambda, \lambda_1, \lambda_2 \in \mathbb{R}$

1. **Associativity:**

$$(\lambda_1 \lambda_2)\mathbf{C} = \lambda_1(\lambda_2 \mathbf{C}); \quad \lambda(\mathbf{BC}) = (\lambda \mathbf{B})\mathbf{C} = \mathbf{B}(\lambda \mathbf{C}) = (\mathbf{BC})\lambda$$

   For $\lambda \in \mathbf{R}$, $\lambda^\top = \lambda$, and hence $(\lambda \mathbf{C})^\top = \mathbf{C}^\top \lambda^\top = \mathbf{C}^\top \lambda = \lambda \mathbf{C}^\top$.

2. **Distributivity:**

$$(\lambda_1 + \lambda_2)\mathbf{A} = \lambda_1 \mathbf{A} + \lambda_2 \mathbf{A}; \; \lambda(\mathbf{A} + \mathbf{B}) = \lambda \mathbf{A} + \lambda \mathbf{B}$$

Compact representation of systems of linear equations

Matrices and operations on matrices are excellent tools for representation of systems of linear equations. For example, the system

$$\begin{cases} 2x_1 + 4x_2 = 1 \\ 4x_1 - 3x_2 = 5 \end{cases}$$

can be represented as the matrix multiplication as

$$\begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

Denoting $\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathrm{b} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$, the above equation can be written in compact form as $\mathbf{Ax} = \mathbf{b}$.

Let's use the linear algebra learned so far to obtain the solution of this system: If $\mathbf{A}$ is invertible, then multiplying the above equation to the left on both sides by $\mathbf{A}^{-1}$ we'd obtain $\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
Check if $\mathbf{A}$ has an inverse: Recall that the inverse exists if $det(\mathbf{A}) \neq 0$.

$$det(\mathbf{A}) = \left| \begin{array}{cc} 2 & 4 \\ 4 & -3 \end{array} \right| = (2)(-3) - (4)(4) = -6 - 16 = -22 \neq 0$$

Next, we actually need to find the inverse: set $\mathbf{A}^{-1} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$, where p_1, p_2, p_3, p_4 must be determined such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2$, that is,

$$\begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 4 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which leads to the system of equations

$$\begin{cases} 2p_1 + 4p_2 = 1 \\ 4p_1 - 3p_2 = 0 \\ 2p_3 + 4p4 = 0 \\ 4p_3 - 3p_4 = 1 \end{cases}$$

whose solution yields $\mathbf{A}^{-1} = \frac{1}{11} \begin{bmatrix} p_1 = 3/2 & p_2 = 2 \\ p_3 = 2 & p_4 = -1 \end{bmatrix}$

Then the solution of the original system is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 3/2 & 2 \\ 2 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} (3/2)(1 + (2)(5) \\ (2)(1) + (-1)(5) \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 23/2 \\ -3 \end{bmatrix}$$