

Deep learning notes

Guillem Tobías Larrauri

January 2025

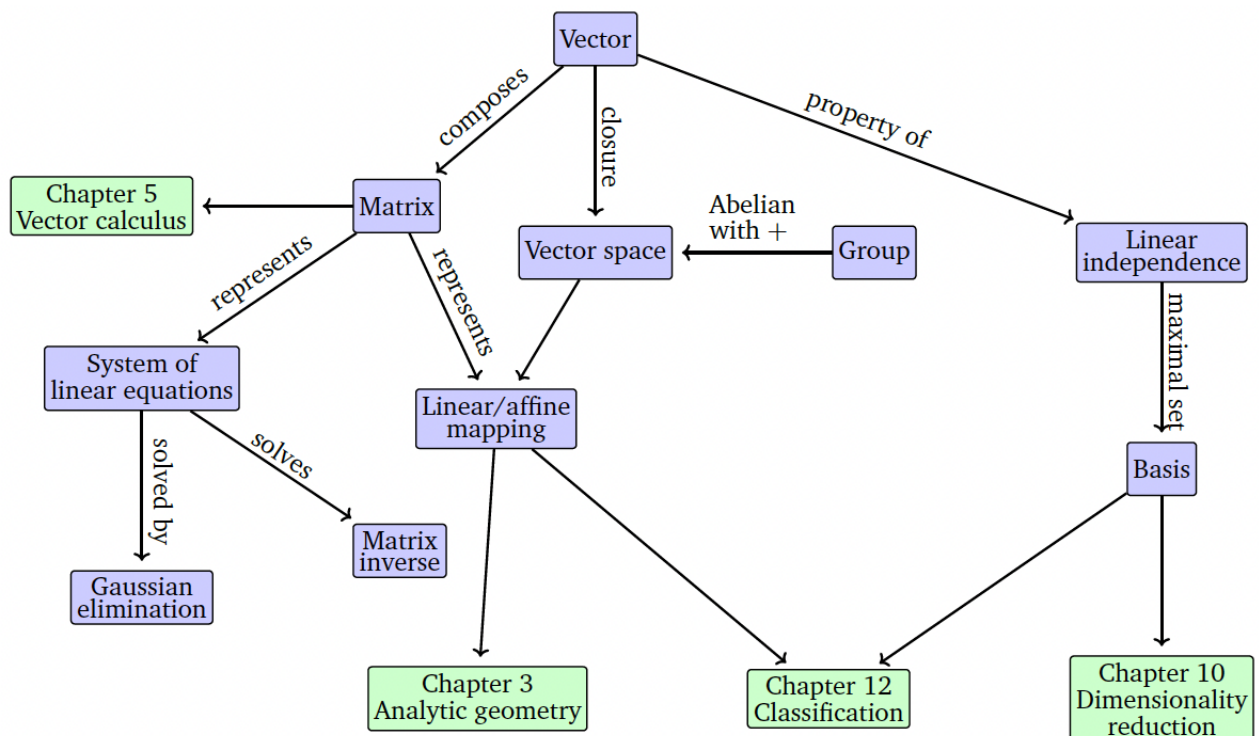
1 Theory

2 Math: From Mathematics for Machine learning

This will be copied over and is needed for ML and also Deep and Graph learning, I will be adding notes on which chapters matter most for where.

2.1 Linear Algebra

An algebra is a set of objects and a set of rules to manipulate these objects. Vectors are objects such that they can be added together and multiplied by a constant and the result is still a vector.



2.1.1 Systems of linear equations

We will mostly ignore for now.

2.1.2 Matrices

A matrix as an $m \times n$ tuple of elements a_{ij} , we define it as \mathbf{A} . We have some important properties of all matrices **TO DO!** $\forall \mathbf{A} \in \mathbb{R}^{m \times n} : \mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A}$

The transpose and the inverse (which not always exists, but when it exists it is unique.) Some important properties of transpose and inverse:

- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A} + \mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1}$
- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $\mathbf{AB}^T = \mathbf{B}^T\mathbf{A}^T$

A matrix is symmetric if it is equal to its transpose and we also have $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} =: \mathbf{A}^{-T}$. For multiplication by a scalar all properties are super intuitive.

2.1.3 Solving systems of linear equations, we skip for now

We can represent compactly via $\mathbf{Ax} = \mathbf{b}$

If the matrix \mathbf{A} is square we can directly compute the solution (if it exists) as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. If the columns of a non square matrix are linearly independent we can compute the following: $\mathbf{Ax} = \mathbf{b} \iff \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b} \iff \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$

2.1.4 Vector spaces

[Add the math if needed.](#) *Groups:* We consider a set \mathcal{G} and an operation $\otimes : \mathcal{G} \times \mathcal{G} \longrightarrow \mathcal{G}$

- Closure: For any x and y in the group, if we apply the operation to x and y the result will still be in the group
- Associativity: For any 3 in the group the parentheses can be changed meaning we do one operation first or second and we get the same result.
- Neutral element: There exists in the group a neutral element such that if we apply the operation to anything in the group the result will be the original thing
- Inverse element: There exists for every element in the group another element such that if we apply the operation we will get the neutral element.

If we have commutative property ($x * y = y * x$) we have **Abelian group**. We then have some examples of what is and what is not a group.

Vector spaces A vector space will now have two operations, an inner operation and an outer operation.

- $+: \mathcal{V} \times \mathcal{V} \longrightarrow \mathcal{V}$
- $*: \mathbb{R} \times \mathcal{V} \longrightarrow \mathcal{V}$

And:

- $(\mathcal{V}, +)$ is an Abelian group.
- We have distributivity with respect to the scalar/outer operation
- We have Associativity with respect to outer operation
- There exists a neutral element with respect to outer operation

Our inner operation will therefore be vector addition and the outer will be multiplication by scalars. Vector multiplication is not defined for all as the dimensions may not be the same.

Examples: I don't understand why $\mathbb{R}^{m \times n}$ is a vector space if the inverse may not exist. AH, inverse with respect to inner operation, meaning find another matrix such that we can get the inverse. And that is always defined by Abelian, meaning commutative under the sum.

Vector subspaces: Subsets of the original space such that when we perform the operations (sum and multiplication) we will never leave the subspace.

We will naturally inherit many of the properties from the larger vector space. However we need to show that U is not empty, that the 0 element is in there and that we have closure with respect to both the outer and inner operations.

Every subspace of \mathbb{R}^n is the solution space of a system of homogeneous linear equations. I don't fully understand the implications of this.

2.1.5 Linear independence

Consider a vector space V and a finite number of vectors (k) . Remember matrices are also vectors in the \mathbb{R}^{nm} vector space.

Every $\mathbf{v} \in V$ of the form $\mathbf{v} = \sum_{i=1}^k \lambda_i \mathbf{x}_i \in V$ where the λ s are real numbers are a linear combination of the initial vectors \mathbf{x} . Notice we apply both operations, the inner operation in the sum and the outer operation. $\mathbf{0}$ can always be written as a linear combination.

The notion of **linear independence** is the following: If from the original vectors we can build a non trivial linear combination that gives us the $\mathbf{0}$ vector the original vectors are linearly dependent. If only setting all λ s to 0 gives us the solution the vectors are linearly independent. *There are a set of practical rules in the book such that we show if they are independent or not, you can also solve the whole system (pain)*

The intuition is that if we drop any of the original vectors we are losing information! [tblueReview](#) page 32 prett cool!

2.1.6 Basis and rank

In a vector space we are interested in a subset of it such that any vector in the larger space can be generated as a linear combination of the vectors in the smaller set. The set of all linear combinations is called the span of a set. If a subset A spans the whole vector space V we say $V = \text{span}(A)$. Generating sets are sets that span the vector space.

A **Basis** is: The minimal generating set. Can also be defined in terms of linear independence, it is the maximal linearly independent set of vectors in V Also uniqueness things, yesyes.

- Every vector space has a basis, however there is not unique basis. All basis have the same number of vectors.
- To find a basis we can see the original vectors, then solve $\sum_{i=1}^k \lambda_i \mathbf{x}_i = \mathbf{0}$ which can be put in matrix form Then find pivot columns and those will be the span.
- Example 2.14 is beautiful [Solving linear systems is useful!!](#)

The **Rank** of a matrix is closely defined: With it we can see the span of the columns and the rows. Columns are m dimensional and so on. [Important remarks for solvable system and existance of an inverse, gp back adn write down!](#)

2.1.7 Linear mappings

[I get lost with this!!! Do a review in the future for now move on.](#) A linear mapping Φ from V to W vector if for any two vectors in the original space and any two real numbers we get $\Phi(\lambda\mathbf{x} + \psi\mathbf{y}) = \lambda\Phi(\mathbf{x}) + \psi\Phi(\mathbf{y})$

We can represent linear mappings as *matrices* and we will see more about this.

Special linear mappings:

- Injective, if we have two vectors equal after transform they were equal in the original space.
- Surjective, we can reach any point in the second vector space from the first vector space using the mapping.
- Bijective if 1 and 2 hold

Bijjective mappings can be undone with the inverse mapping, so this guarantees existence or whatever. Isomorphic is if linear and bijective from V to W .

We have a theorem that says that V and W are isomorphic iff they have the same dimension. This justifies treating $\mathbb{R}^{m \times n} = \mathbb{R}^{mn}$ we can unstack the matrix and not lose any information. Because there exists a bijective linear mapping that goes from one to the other.

There are some extra properties that I can write down in p 38.

Matrix representation of linear mappings: Bottom line is that linear mappings can be represented as matrices.

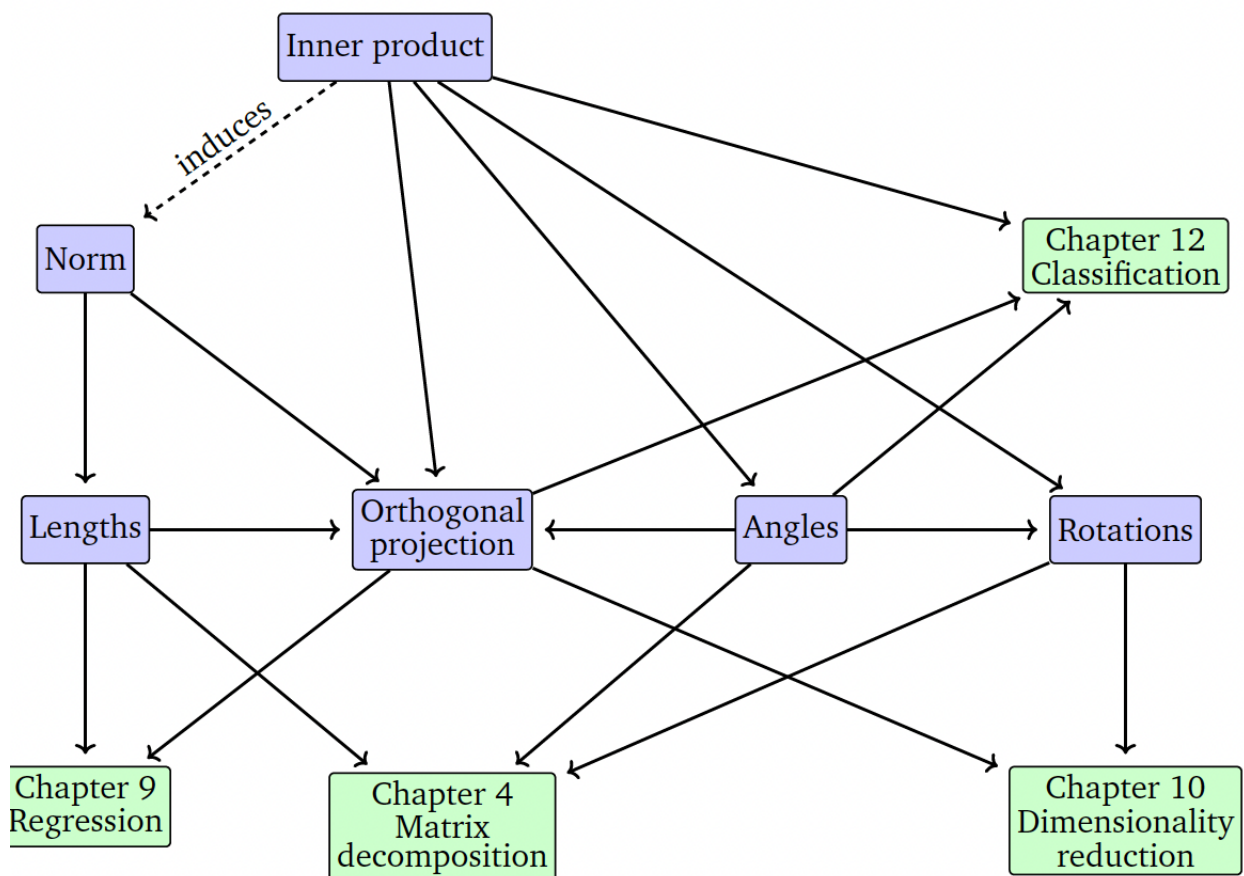
Any n dimensional vector space is isomorphic to \mathbb{R}^n We will then define $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ as an ordered basis.

Then for any vector space V with its ordered basis B , we can define any vector inside the space as $\sum_{i=1}^n \alpha_i \mathbf{b}_i$ and this representation is unique, and the alphas are the coordinate representation of the vector x with basis B . [Example 2.20 explains this but I still don't fully understand the uses. I don't fully understand definition 2.19](#) Some transformation matrix to go from one ordered basis to other coordinates in W . And if we have coordinates in the other space we can uniquely identify the vector by 2.18?! [I think this is it more or less.](#)

2.1.8 Exercises and complete above

2.2 Analytic geometry

Goes into geometric interpretation for linear algebra, lengths and distances between two vectors inner products and norms and finally orthogonal projections.



2.2.1 Norms

A norm is defined on a vector space and goes from the vector space to a real number. It defines the length of a vector. To be a proper norm the following need to hold for any two vectors in the space and any real number lambda:

- $\|\lambda \mathbf{x}\| = \text{abs}(\lambda) \|\mathbf{x}\|$
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- Positive definite meaning larger or equal than 0 and 0 for the 0 vector.

We usually use the Euclidean norm $\sqrt{\mathbf{x}^T \mathbf{x}}$ which computes the euclidean distance from the origin.

2.2.2 Inner products

Allow introduction of geometrical concepts like length angle and distance between two vectors. The simplest is the dot product.

The general definition of a inner product is as a positive definite (no neg distances), symmetric (distance a to b is same as b to a) and bilinear (why not) such that it maps from $V \times V \rightarrow \mathbb{R}$

3 Tensorflow