

# Mathematics for Machine Learning - Notes

September 19, 2023

## Chapter 2

- Vector space is a bunch of "things" with an addition function and scalar multiplication (over some field), with closure under each.
- Dimension is the smallest number of "vectors" in this space that we need to be able to generate the entire space with linear combinations thereof. We call any of these sets of generating vectors a "basis".
- In maths we don't just care about things - we care about maps between things. We can define a linear map as something that "respects linearity" - i.e. we map from one vector space to another, both over the same field, with a map that distributes over addition and commutes with scalar multiplication.
- Suppose we have a linear map  $\Phi$  from space  $A$  of dimension  $m$  to space  $B$  of dimension  $n$ . Then if we pick any basis of space  $A$ , then the action of  $\Phi$  on these  $m$  basis vectors uniquely defines  $\Phi$ . Each of these is mapped to a unique linear combination of the image basis. Therefore with respect to a domain and image basis, we can completely define a linear transformation with just  $m * n$  variables. This is how we arrive at the concept of a matrix.
- Note that a matrix is the realisation of an abstract linear map into a specific choice of domain and image basis. However, when it comes to commonly used vector spaces like  $\mathbb{R}^n$ , there is a canonical basis we often use.
- Matrix addition doesn't really have a useful geometric interpretation, but matrix multiplication does. We can multiply matrices together to "pre-jack" our calculations and reduce the amount of work we have to do.
- We can also use matrices to solve systems or linear equations, but possibly a better way to look at this is "what things can we put into this transformation to get the output we want". We will see how this becomes useful later - ML is all about tailoring a system to get specific outputs under certain circumstances.

- I like to describe the general solution to a matrix equation (or linear system, if you prefer) as a specific non-homogeneous solution plus the general solution to the homogenous problem.
- In general the best way to do this is via row reductions. We are allowed to switch rows, multiply through by non-zero constants, and add any multiple of a row to a distinct row. Note that all of these operations are completely reversible. This guarantees that the solution set is unaffected.
- A matrix is in reduced echelon row form when all pivots are 1s, each non-empty row is shorter than the one above it, and all other elements in a pivot column are zero
- We can use row reductions to put things in this form. This way we have free choice over a bunch of different variables (columns with no pivot) and can easily just "read off" the dependent ones (columns with a pivot since they appear in exactly one equation, with a coefficient of 1)
- Previously we used row reduction to solve a system of linear equations, but it can also be used to determine a matrix inverse from  $AA^{-1} = I$ . As we do row operations on  $A$  to convert it into reduced echelon row form, the final result will be the identity matrix, and the matrix on the right will be warped into the actual numerical  $A^{-1}$ .
- In reality there are better practical ways to solve linear systems and Gaussian elimination isn't used.
- This chapter has some other information but I don't really care. The above are the important things I didn't know all that well, expressed in a good way.