

Vector Spaces

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 \\ 1 & 3 & 9 & 27 & 81 \\ 1 & 4 & 16 & 64 & 256 \\ 1 & 5 & 25 & 125 & 625 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ -3 \\ 1 \end{pmatrix}$$

$$M\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\mathbf{b} = M^{-1} \mathbf{x}$$

$$\text{Tr}(X^{-1}A) = -X^{-1}AX^{-1}$$

Vectors, metric spaces, norms

Matrices, Vectors and All That

- The language of machine learning is mathematics
- Sometimes we draw pretty pictures to explain the mathematics
- Much of the mathematics we will use involves vectors, matrices and functions
- You need to master the language of mathematics, otherwise you won't understand the algorithms
- I'm going to spend this lecture and the next revising the mathematics you need to know (but I'm going use a slightly posher language than you are probably used to)

Vectors

- We often work with objects with many components (features)
- To help handle this we will use vector notation

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

- We represent vectors by bold symbols
- All our vectors are column vectors by default
- We treat them as $n \times 1$ matrix

- We write row vectors as transposes of column vectors

$$\mathbf{y}^T = (y_1, y_2, \dots, y_n)$$

Vector Space

- A vector space, \mathcal{V} , is a set of vectors which satisfies
 - if $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ then $a\mathbf{v} \in \mathcal{V}$ and $\mathbf{v} + \mathbf{w} \in \mathcal{V}$ (closure)
 - $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ (commutativity of addition)
 - $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (associativity of addition)
 - $\mathbf{v} + \mathbf{0} = \mathbf{v}$ (existence of additive identity $\mathbf{0}$)
 - $1\mathbf{v} = \mathbf{v}$ (existence of multiplicative identity 1)
 - $a(b\mathbf{v}) = (ab)\mathbf{v}$ (distributive properties)
 - $a(\mathbf{v} + \mathbf{w}) = a\mathbf{v} + a\mathbf{w}$
 - $(a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

(You don't need to remember these)

- Just from these properties we can deduce other properties

- Vector Spaces
- Metrics (distances)
- Norms

$$M\mathbf{x} = \mathbf{b}$$

$$M\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\mathbf{b} = M^{-1} \mathbf{x}$$

Scalars (Fields)

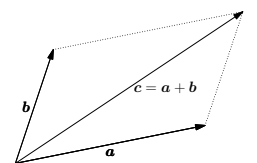
- Vector spaces involve fields (numbers)—aka scalars
- These are quantities we can add together ($a + b$) and multiply together ($a \times b$)
- Formally they form an Abelian group under addition with an identity 0 and excluding 0 an Abelian group under multiplication and they are distributive

$$a \times (b + c) = a \times b + a \times c$$

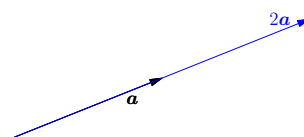
- Although this sounds rather daunting don't panic. They behave like numbers. The field might be integers, rational numbers, reals, complex numbers or something a bit more exotic—but we will almost always consider reals

Basic Vector Operations

- The basic vector operations are adding

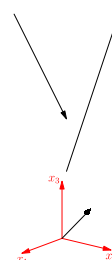


- multiplying by a scalar (a number)



$$\mathbb{R}^n$$

- When we first learn about vectors we think of them arrows in 3-D space
- If we centre them all at the origin then there is a one-to-one correspondence between vectors and points in space
- We call this vector space \mathbb{R}^3
- Any set of quantities $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ which satisfy the axioms above form a vector space \mathbb{R}^n
- Of course, we can't so easily draw pictures of high-dimensional vectors



- Any set of object that satisfies the axioms of a vector spacer are **vectors** —not just $v \in \mathbb{R}^n$
- Matrices satisfy all the conditions of a vector space
- Infinite sequences form a vector space
- Functions form a vector space
 - Let $C(a,b)$ be the set of functions defined on the interval $[a,b]$
 - Note that if $f(x), g(x) \in C(a,b)$ then $af(x) \in C(a,b)$ and $f(x) + g(x) \in C(a,b)$
- Bounded vectors in \mathbb{R}^n **don't** form a vector space

Metrics

- Vector spaces become more interesting if we have a notion of distance
- We say $d(x,y)$ is a **proper distance** or **metric** if
 - $d(x,y) \geq 0$ (non-negativity)
 - $d(x,y) = 0$ iff $x = y$ (identity of indiscernibles)
 - $d(x,y) = d(y,x)$ (symmetry)
 - $d(x,y) \leq d(x,z) + d(z,y)$ (triangular inequality)
- There are typically many possible distances (e.g. Euclidean distance, Manhattan distance, etc.)
- Often one or more condition isn't satisfied then we have a **pseudo-metric**

Lipschitz Function

- One way to characterise well behaved function, $f(x)$ is if there exists a number $K < \infty$ such that for all x and y

$$d(f(x), f(y)) \leq K d(x,y)$$
- This is known as a **Lipschitz condition** and the function is said to be K -Lipschitz or Lipschitz continuous
- Note that such functions cannot have any jumps (i.e. they are continuous)
- The size of K measures the limit on the amplifying effect of the function

Outline

- Vector Spaces
- Metrics (distances)
- Norms**

$$Mx=b$$

$$Mv_i = \lambda_i v_i$$

$$b=M \cdot$$

- Vector Spaces
- Metrics (distances)**
- Norms

Mappings and Functions

- A function defines a mapping from one vector space to another (although the spaces might be the same) e.g.

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

(f maps the reals onto reals, i.e. $f(x)$ takes a real x and gives you a new real number $y = f(x)$)

- We are often interested in functions that behave nicely
- E.g. They are continuous

Contractive Mappings

- An interesting class of function are those for which $K < 1$
- These are said to be contractive mappings
- A famous theorem that applies to contractive mappings is the Banach fixed-point theorem which says there exists a unique fixed point such that $f(x) = x$
- This is used for example in showing that various algorithms will converge

Norms

- Vector spaces are even more interesting with a notion of length
- Norms** provide some measure of the size of a vector
- To formalise this we define the **norm** of an object v as $\|v\|$ satisfying
 - $\|v\| > 0$ if $v \neq 0$ (non-negativity)
 - $\|av\| = a\|v\|$ (linearity)
 - $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality)
- When some criteria aren't satisfied we have a **pseudo-norms**
- Norms provide a metric $d(x,y) = \|x - y\|$ (they are metric spaces)

Vector Norms

- The familiar vector norm is the (Euclidean) two norm

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

- Other norms exist, such as the p -norm ($p \geq 1$)

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}$$

- Special cases include the 1-norm and the infinite norm

$$\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i| \quad \|\mathbf{v}\|_\infty = \max_i |v_i|$$

- The 0-norm is a pseudo-norm as it does not satisfy condition 2

$$\|\mathbf{v}\|_0 = \text{number of non-zero components}$$

Matrix Norms

- We can define norms for other objects
- The norm of a matrix encodes how large the mapping is
- The Frobenius norm is defined by

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2}$$

- Many other norms exist including 1-norm, max-norm, etc.
- For square matrices, some, but not all, norms satisfy the inequality

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \times \|\mathbf{B}\|$$

Compatible Norms

- A vector and matrix norm are said to be compatible if

$$\|\mathbf{M}\mathbf{v}\|_b \leq \|\mathbf{M}\|_a \times \|\mathbf{v}\|_b$$

(Spectral and Euclidean norms are compatible)

- Norms provide quick ways to bound the maximum growth of a vector under a mapping induced by the matrix
- We will see that a measure of the sensitivity of a mapping is in terms of the ratio of its maximum effect to its minimum effect on a vector
- This is known as the **conditioning**, given by $\|\mathbf{M}\| \times \|\mathbf{M}^{-1}\|$

Function Norms

- Functions can also have norms, for example, if $f(x)$ is defined in some interval \mathcal{I}

$$\|f\|_{L_2} = \sqrt{\int_{x \in \mathcal{I}} f^2(x) dx}$$

- The L_2 vector space is the set of function where $\|f\|_{L_2} < \infty$
- The L_1 -norm is given by $\|f\|_{L_1} = \int_{x \in \mathcal{I}} |f(x)| dx$
- The infinite-norm is given by $\|f\|_\infty = \max_{x \in \mathcal{I}} |f(x)|$

Why Should You Care?

- Deep learning involves multiply the input (which we can think of as a vector \mathbf{x}) by many layers
- In CNNs we have convolutional layers and dense layers
- The effect of applying these layers can be represented by a matrix multiplication $\mathbf{x}_n = \mathbf{L}_n \mathbf{x}_{n-1}$
- We also do other things like applying ReLU's or pooling that changes the magnitude, \mathbf{x}_n , of our representation
- If you are developing new architectures you want $\|\mathbf{x}_n\|$ neither to blow up or vanish
- This can be controlled by carefully choosing $\|\mathbf{L}_n\|$

Summary

- Vector spaces with a distance (metric spaces) and vector spaces with a norm (normed vector spaces) are interesting objects
- They allow you to define a topology (open/closed sets, etc.)
- You can build up ideas about connectedness, continuity, contractive maps, fixed-point theorems, . . .
- For the most part we are going to consider an even more powerful vector space that has an inner-product defined