Advanced Machine Learning Subsidary Notes

Lecture 5: Vector Spaces

Adam Prügel-Bennett

January 28, 2021

1 Keywords

- Vectors, vector spaces, operators

2 Main Points

2.1 Vector Spaces

- Any set of objects with addition between members of the set and scalar multiplication forms a vector space if they satisfies 8 axioms
- Most of these axioms arise naturally if addition and scale multiplication behave normally
- The only additional axiom is closure
- Normal vectors, matrices and functions all form vector spaces

2.2 Distances

- A *proper distance* or *metric* between objects in a vector space satisfies 4 conditions
  1. $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ (non-negativity)
  2. $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} = \boldsymbol{y}$ (identity of indiscernibles)
  3. $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$ (symmetry)
  4. $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$ (triangular inequality)
- You can define different distances for the same set of objects
- Often we use *pseudo-metrics* that breaks one or other of the conditions

2.3 Norms

- Norms provide a measure of the size of vector
- They satisfy three conditions
  1. $\|\boldsymbol{v}\| > 0$ if $\boldsymbol{v} \neq \boldsymbol{0}$ (non-negativity)
  2. $\|a\,\boldsymbol{v}\| = a\|\boldsymbol{v}\|$ (linearity)
  3. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ (triangular inequality)
- Again if not all of these conditions are true we have *pseudo-norms*
- Norms provide a metric $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$
- We will meet norms very often in this course

- **Vector Norms**
  - There are a large number of norms for normal vectors that people use
    1. Euclidean or 2-norm: $\|\boldsymbol{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$
    2. p-norm: $\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^n |v_i|^p\right)^{1/p}$
    3. 1-norm: $\|\boldsymbol{v}\|_1 = \sum\limits_{i=1}^n |v_i|$
    4. ∞-norm or max-norm: $\|\boldsymbol{v}\|_\infty = \max_i |v_i|$
  - Note the 1-norm, 2-norm and ∞-norm are all p-norms with different p
  - The 0-norm counts the number of non-zero components (it is a pseudo-norm as it is not linear)

- **Matrix Norms**
  - Matrices also have norm
    1. The Frobenius norm is $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2}$
    2. Also have 1-norm, max-norm, Hilbert-norm (the maximum absolute eigenvalue), nuclear-norm, etc.
  - Note that the determinant is not a norm because it can be negative and is not linear
  - Many of the commonly used matrix norms satisfy

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

  - This is really useful because we can quickly bound norms of products of matrices
  - Many matrix and vector norms are compatible

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

  - E.g. Frobenius and Euclidean norms are compatible
  - One of the main uses of matrix norms is to understand by how much it can potentially increase the size of a vector

- **Function Norms**
  - The most common function norms are
    1. The L_2-norm

$$\|f\|_{L_2} = \sqrt{\int_{\boldsymbol{x}\in\mathcal{R}} f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}$$

       where $\mathcal{R}$ is the region over which the function is define
    2. The L_1-norm

$$\|f\|_{L_1} = \int_{\boldsymbol{x}\in\mathcal{R}} |f(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}$$

    3. The ∞ or max-norm

$$\|f\|_\infty = \max_{\boldsymbol{x}\in\mathcal{R}} f(\boldsymbol{x})$$

  - Function norms are also used to define vector spaces
    1. The L_2 vector space is the set of functions such that all functions satisfy $\|f\|_{L_2} < \infty$
    2. The L_1 vector space is the set of functions such that all functions satisfy $\|f\|_{L_1} < \infty$
  - In these vector spaces we only consider functions that measurable in the sense that $\|f\| > 0$ for any non-zero function

2.4 Inner Products

- For some vector spaces we can sometimes define a *inner product*
- Inner products are scalars associated with two elements in a vector space
- They are generally denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$
- For normal vectors the standard inner product is the dot-product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y} = \sum_{i=1}^{n} x_i \, y_i$$

- We can define an inner product between functions

$$\langle f, g \rangle = \int_{x \in \mathcal{I}} f(x) \, g(x) \, \mathrm{d}x$$

- For lots of vector spaces we don't bother defining inner products (e.g. we don't often do this matrices)
- Inner products allow us to define the notion of similarity

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{x}\| \, \cos(\theta)$$
$$\langle f(x), g(x) \rangle = \|f(x)\| \, \|g(x)\| \, \cos(\theta)$$

- $\cos(\theta)$ can be seen as a measure of the correlation between vectors (or functions)

2.5 Coordinates or Basis Vectors

- Any set of vectors that span the entire vector space can be considered a set of basis vectors or coordinates
- If our bases are linearly independent then we have a set of non-degenerate basis function where each vector is unique
- The most convenient set of basis vectors are those where the bases are normalised and orthogonal $\langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle = \delta_{ij}$
- **Basis Functions**
  - For a function space we can consider a set of basis functions
  - A familiar set of functions define on the interval $[0, 2\,\pi]$ are the Fourier functions

$$\{1, \cos(\theta), \sin(\theta), \cos(2\,\theta), \sin(2\,\theta), \cdots\}$$

  - This basis set is orthogonal as for any two components $\langle b_i(\theta), \, b_j(\theta) \rangle = \delta_{ij}$
  - There are many orthogonal polynomials that have similar properties
  - Given an orthogonal set of functions $\{b_i(\boldsymbol{x})\}$ we can decompose a function $f(\boldsymbol{x})$ as a (infinite) vector $\boldsymbol{f}$ with components $f_i = \langle f(\boldsymbol{x}), b_i(\boldsymbol{x}) \rangle$
  - This allows us to represent any function as a point in an infinite-dimensional space

2.6 Operators

- Operators transform elements of a vector space
- Consider the transformation or operator $\mathcal{T} : \mathcal{V} \to \mathcal{V}'$
- This says that $\mathcal{T}$ maps some object $\boldsymbol{x} \in \mathcal{V}$ to an object $\boldsymbol{y} = \mathcal{T}[\boldsymbol{x}]$ in a new vector space $\mathcal{V}'$
- **Linear Operators**
  - Linear operators satisfy the two conditions
    1. $\mathcal{T}[a\,\boldsymbol{x}] = a\,\mathcal{T}[\boldsymbol{x}]$
    2. $\mathcal{T}[\boldsymbol{x}+\boldsymbol{y}] = \mathcal{T}[\boldsymbol{x}] + \mathcal{T}[\boldsymbol{y}]$
  - Linear operators are really important because we can understand them
  - For normal vectors the most general linear operation is

    $$\mathcal{T}[\boldsymbol{x}] = \mathbf{M}\,\boldsymbol{x}$$

    where $\mathbf{M}$ is a matrix
  - For functions the most general linear operator is a kernel function

    $$g(\boldsymbol{x}) = \mathcal{T}[f(\boldsymbol{x})] = \int K(\boldsymbol{x},\boldsymbol{y})\,f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

    * Kernels appear in SVMs, SVRs, kernel-PCA, Gaussian Processes
- Often we are interested in operators that map vectors in a vector space to new vectors in the same vector space
  - $\mathcal{T} : \mathcal{V} \to \mathcal{V}$
  - The most general linear mapping for normal vectors that does this are square matrices