Advanced Machine Learning Subsidary Notes

Lecture 5: Vector Spaces

Adam Prügel-Bennett

January 17, 2024

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
- Consider a function (mapping) $f : \mathbb{R}\mathbb{R}$ then a useful quantification of how big a change a mapping can produce is given by the Lipschitz condition

$$d(f(x), f(y)) \leq K\, d(x, y)$$

  This is useful if for all x and y there exists some fixed K
- Lipschitz functions are continuous (have no jumps)
- If $K < 1$ the mapping is said to be a contractive mapping

2.3 Norms

- Norms provide a measure of the size of a 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