Advanced Machine Learning Subsidary Notes

Lecture 8: Inner-Product Spaces

Adam Prügel-Bennett

February 26, 2024

1 Keywords

- Inner-product, operators

2 Main Points

- For some vector spaces we can define an *inner product* between pairs of vectors
- Inner products are scalars associated with two elements in a vector space
- They are generally denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$
- To be an inner-product requires satisfying 5 axioms
  1. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x} \in \mathcal{V}$
  2. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$
  3. $\langle \alpha\, \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \, \langle \boldsymbol{x}, \boldsymbol{y} \rangle$
  4. $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$
  5. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
- The inner-product induces a norm: $\|x\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$
- For normal vectors (i.e. $\boldsymbol{x} \in \mathbb{R}^n\}$) 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 matrices we can define the Frobenius inner-product

$$\langle \mathbf{A}, \mathbf{B} \rangle_F = \mathrm{Tr} \, \mathbf{A}^{\mathsf{T}} \, \mathbf{B} = \sum_{i,j} A_{ij} \, B_{ij}$$

  which defines the Frobenius norm

$$\|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle_F} = \sqrt{\sum_{i,j} A_{ij}^2}$$

  The Frobenius norm is not a compatible norm.

- A really important inequality is the *Cauchy-Schwarz* inequality

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leq \langle \boldsymbol{x}, \boldsymbol{y} \rangle \; \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 \; \|\boldsymbol{y}\|^2$$

- 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)
- Because of Cauchy-Schwarz $\langle \boldsymbol{x}, \boldsymbol{y} \rangle / (\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|)$ lies between -1 and 1 (so that we can represent this quantity by the cosine of an angle)

2.1 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

2.2 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

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

    where 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

2.3 Matrices and Mappings

- Matrices, $\mathbf{M}$ are linear maps from one point $\boldsymbol{x}$ to another $\boldsymbol{y} = \mathbf{M}\boldsymbol{x}$
- The product of a matrix $\mathbf{C} = \mathbf{A}\mathbf{B}$ corresponds to applying the mapping $\mathbf{B}$ followed by $\mathbf{A}$
- For most matrices $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$ (we say matrix multiplication does not commute)
- There are pairs of matrices that do commute, but they need to share a special structure for this to happen
- Matrix multiplication is associative. That is, $(\mathbf{A}\,\mathbf{B})\mathbf{C} = \mathbf{A}(\mathbf{B}\,\mathbf{C})$