Lecture 1c. Linear Algebra Review

COMP90051 Statistical Machine Learning

Sem2 2020
Lecturer: Ben Rubinstein

This lecture

- About COMP90051
- Review: Probability theory
- **Review: Linear algebra**
- Review: Sequences and limits

Vectors

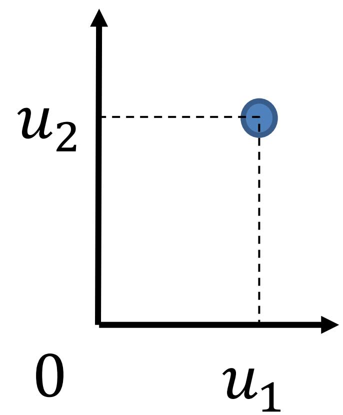
Link between geometric and algebraic interpretation of ML methods

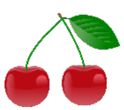
What are vectors?

Suppose $\boldsymbol{u} = [u_1, u_2]'$. What does $\boldsymbol{u}$ really represent?

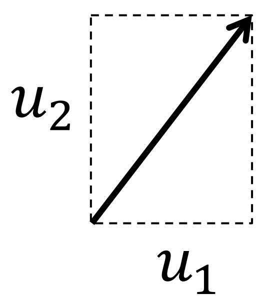
Ordered set of numbers $\{u_1, u_2\}$

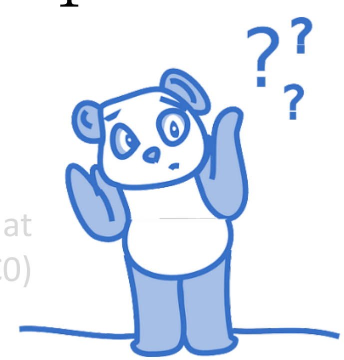
Cartesian coordinates of a point





A direction





art: OpenClipartVectors at pixabay.com (CC0)

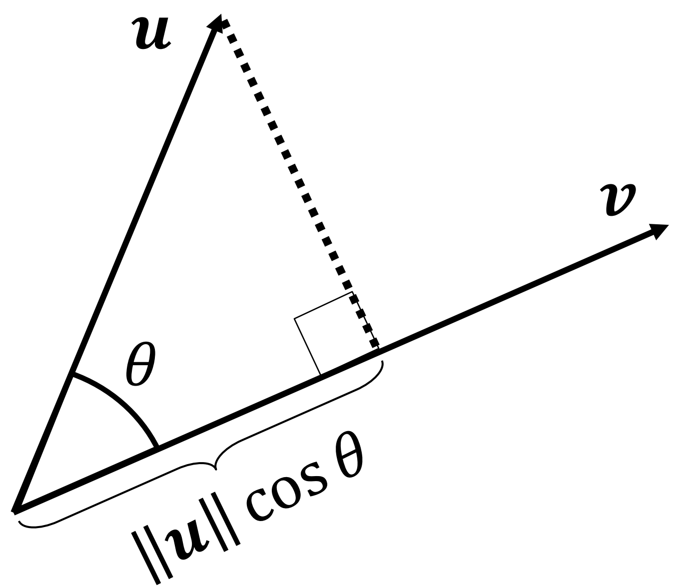
Dot product: Algebraic definition

- Given two m-dimensional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, their dot product is $\boldsymbol{u} \cdot \boldsymbol{v} \equiv \boldsymbol{u}'\boldsymbol{v} \equiv \sum_{i=1}^{m} u_i v_i$
  - E.g., weighted sum of terms is a dot product $\boldsymbol{x}'\boldsymbol{w}$
- If k is a scalar, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are vectors then

$$(k\boldsymbol{a})'\boldsymbol{b} = k(\boldsymbol{a}'\boldsymbol{b}) = \boldsymbol{a}'(k\boldsymbol{b})$$
$$\boldsymbol{a}'(\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a}'\boldsymbol{b} + \boldsymbol{a}'\boldsymbol{c}$$

Dot product: Geometric definition

- Given two m-dimensional Euclidean vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, their dot product is $\boldsymbol{u} \cdot \boldsymbol{v} \equiv \boldsymbol{u}'\boldsymbol{v} \equiv \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta$
  - $\|\boldsymbol{u}\|, \|\boldsymbol{v}\|$ are L_2 *norms* for $\boldsymbol{u}, \boldsymbol{v}$ also written as $\|\boldsymbol{u}\|_2$
  - θ is the angle between the vectors



The *scalar projection* of $\boldsymbol{u}$ onto $\boldsymbol{v}$ is given by

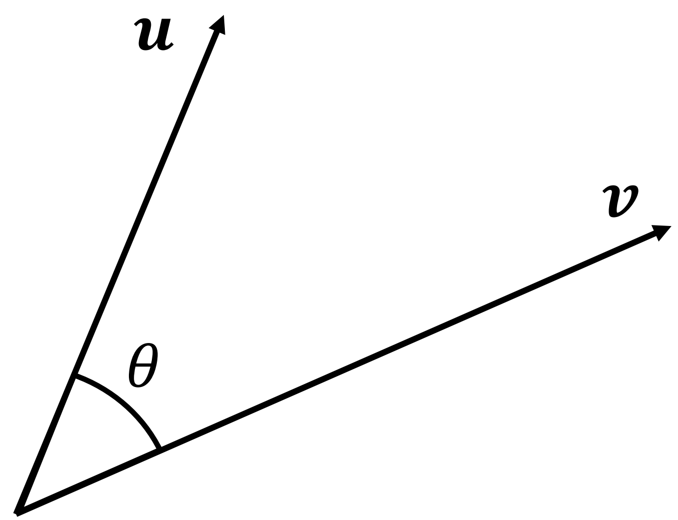
$$u_{\boldsymbol{v}} = \|\boldsymbol{u}\| \cos\theta$$

Thus dot product is

$$\boldsymbol{u}'\boldsymbol{v} = u_{\boldsymbol{v}}\|\boldsymbol{v}\| = v_{\boldsymbol{u}}\|\boldsymbol{u}\|$$

Geometric properties of the dot product

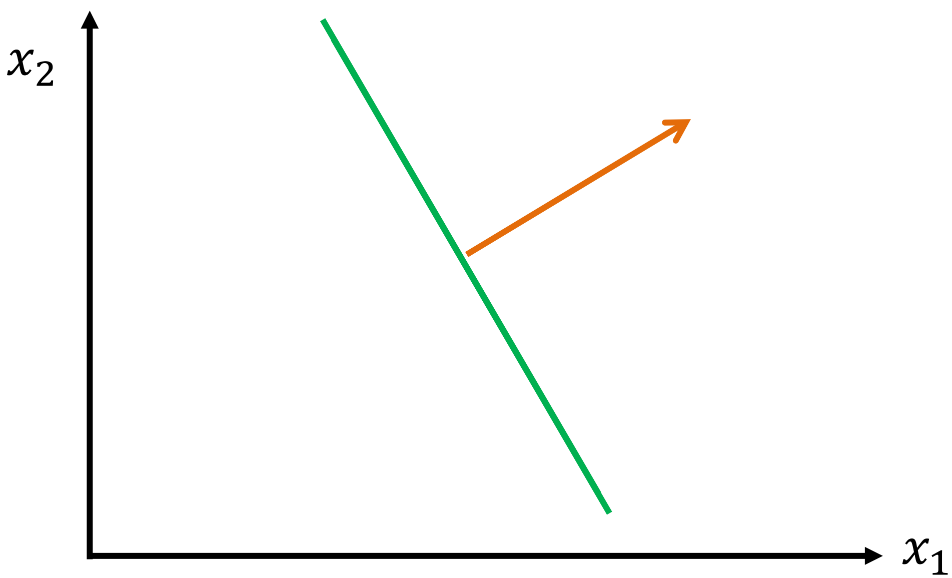
- If the two vectors are orthogonal then $\boldsymbol{u}'\boldsymbol{v} = 0$
- If the two vectors are parallel then $\boldsymbol{u}'\boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\|$, if they are anti-parallel then $\boldsymbol{u}'\boldsymbol{v} = -\|\boldsymbol{u}\|\|\boldsymbol{v}\|$
- $\boldsymbol{u}'\boldsymbol{u} = \|\boldsymbol{u}\|^2$, so $\|\boldsymbol{u}\| = \sqrt{u_1^2 + \cdots + u_m^2}$ defines the Euclidean vector length



Hyperplanes and normal vectors

- A ***hyperplane*** defined by parameters $\boldsymbol{w}$ and b is a set of points $\boldsymbol{x}$ that satisfy $\boldsymbol{x}'\boldsymbol{w} + b = 0$

- In 2D, a hyperplane is a line: a line is a set of points that satisfy $w_1x_1 + w_2x_2 + b = 0$



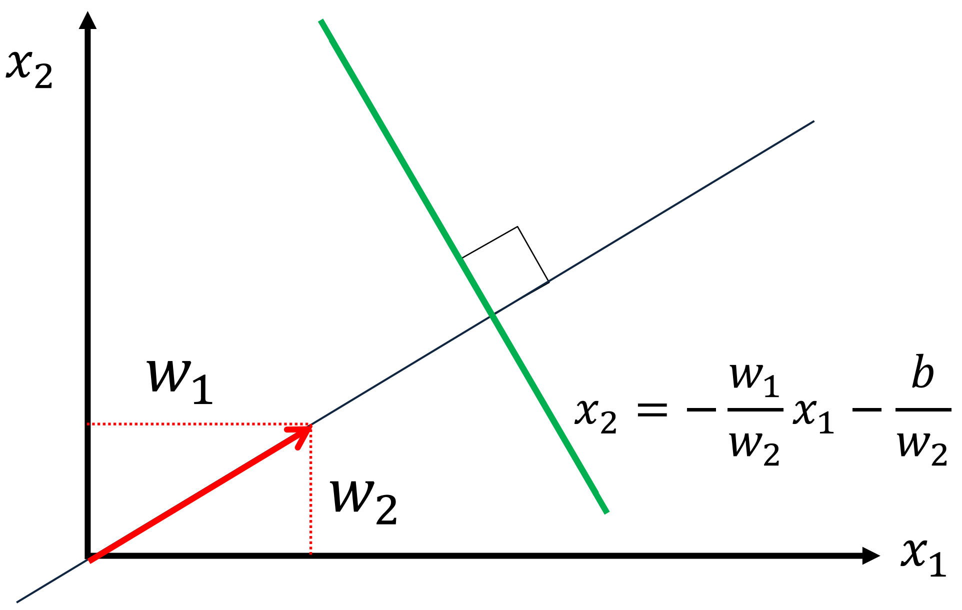
A ***normal vector*** for a hyperplane is a vector perpendicular to that hyperplane

Hyperplanes and normal vectors

- Consider a hyperplane defined by parameters $\boldsymbol{w}$ and b. Note that $\boldsymbol{w}$ is itself a vector
- Lemma: Vector $\boldsymbol{w}$ is normal to the hyperplane
- Proof sketch:
  * Choose any two points $\boldsymbol{u}$ and $\boldsymbol{v}$ on the hyperplane. Note that vector $(\boldsymbol{u} - \boldsymbol{v})$ lies on the hyperplane
  * Consider dot product $(\boldsymbol{u} - \boldsymbol{v})'\boldsymbol{w} = \boldsymbol{u}'\boldsymbol{w} - \boldsymbol{v}'\boldsymbol{w}$
  $$= (\boldsymbol{u}'\boldsymbol{w} + b) - (\boldsymbol{v}'\boldsymbol{w} + b) = 0$$
  * Thus $(\boldsymbol{u} - \boldsymbol{v})$ lies on the hyperplane, but is perpendicular to $\boldsymbol{w}$, and so $\boldsymbol{w}$ is a vector normal

Example in 2D

- Consider a line defined by w_1, w_2 and b
- Vector $\boldsymbol{w} = [w_1, w_2]'$ is a normal vector



L_1 and L_2 norms

- Throughout the subject we will often encounter norms that are functions $\mathbb{R}^n \to \mathbb{R}$ of a particular form
  - Intuitively, norms measure lengths of vectors in some sense
  - Often component of objectives or stopping criteria in optimisation-for-ML
- More specifically, we will often use the L_2 norm (*aka* Euclidean distance)

$$\|\boldsymbol{a}\| = \|\boldsymbol{a}\|_2 \equiv \sqrt{a_1^2 + \cdots + a_n^2}$$

- And also the L_1 norm (*aka* absolute norm or Manhattan distance)

$$\|\boldsymbol{a}\|_1 \equiv |a_1| + \cdots + |a_n|$$

Vector Spaces and Bases

Useful in interpreting matrices and some algorithms like PCA

Linear combinations, Independence

- For formal definition of vector spaces:
  https://en.wikipedia.org/wiki/Vector_space#Definition
- A linear combination of vectors $v_1, \ldots, v_k \in V$ some vector space, is a new vector $\sum_{i=1}^{k} a_i v_i$ for some scalars $a_1, \ldots, a_k$
- A set of vectors $\{v_1, \ldots, v_k\} \subseteq V$ is called linearly dependent if one element v_j can be written as a linear combination of the other elements
- A set that isn't linearly dependent is linearly independent

Spans, Bases

- The span of vectors $v_1, \ldots, v_k \in V$ is the set of all obtainable linear combinations (ranging over all scalar coefficients) of the vectors
- A set of vectors $\{v_1, \ldots, v_k\} \subseteq V$ is called a basis for a vector subspace $V' \subseteq V$ if
  1. The set is linearly independent; and
  2. Every $\mathrm{v} \in V'$ is a linear combination of the set.
- An orthonormal basis is a basis in which each
  1. Pair of basis vectors are orthogonal (zero dot prod); and
  2. Basis vector has norm equal to 1.

Matrices

Some useful facts for ML

Basic matrices

- See more: https://en.wikipedia.org/wiki/Matrix_(mathematics)
  - Including matrix-matrix and matrix-vector products
- A rectangular array, often denoted by upper-case, with two indices first for row, second for column
- Square matrix has equal dimensions (numbers of rows and columns)
- Matrix transpose **A**’ or $\mathbf{A}^{T}$ of *m* by *n* matrix **A** is an *n* by *m* matrix with entries $A'_{ij}=A_{ji}$
- A square matrix **A** with **A**=**A**’ is called symmetric
- The (square) identity matrix **I** has 1 on the diagonal, 0 off-diagonal
- Matrix inverse $\mathbf{A}^{-1}$ of square matrix **A** (if it exists) satisfies $\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$

Matrix eigenspectrum

- Scalar, vector pair (λ, v) are called an eigenvalue-eigenvector pair of a square matrix **A** if $\mathbf{Av} = \lambda\mathbf{v}$
  - Intuition: matrix **A** doesn't rotate **v** it just stretches it
  - Intuition: the eigenvalue represents stretching factor
- In general eigenvalues may be zero, negative or even complex (imaginary) – we'll only encounter reals

Spectra of common matrices

- Eigenvalues of symmetric matrices are always real (no imaginary component)
- A matrix with linear dependent columns has some zero eigenvalues (called rank deficient) → no matrix inverse exists

Positive (semi)definite matrices

- A symmetric square matrix **A** is called positive semidefinite if for all vectors **v** we have $\mathbf{v}'\mathbf{A}\mathbf{v} \geq 0$.
  - Then **A** has non-negative eigenvalues
  - For example, any $\mathbf{A} = \mathbf{X}'\mathbf{X}$ since: $\mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = \|\mathbf{X}\mathbf{v}\|^2 \geq 0$
- Further if $\mathbf{v}'\mathbf{A}\mathbf{v} > 0$ holds as a strict inequality then **A** is called positive definite
  - Then **A** has (strictly) positive eigenvalues

Summary

- Vectors: Vector spaces, dot products, independence, hyperplanes
- Matrices: Eigenvalues, positive semidefinite matrices

Next time: Sequences and limits review/primer