Vectors

Lecture recap

- We represent *vectors* using two styles of notation

$$\begin{bmatrix} -1.2 \\ 0.0 \\ 3.6 \\ -7.2 \end{bmatrix} \quad \text{or} \quad (-1.2, 0.0, 3.6, -7.2).$$

- *Block or stacked vectors* can be produced by *concatenating* two or more vectors, as in

$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix} \quad \text{or} \quad a = (b, c, d),$$

where a, b, c and d are vectors. For example, if $b = (2, 8)$ and $c = (3, -1, 7)$, then

$$a = (b, c) = (2, 8, 3, -1, 7).$$

- *Subvectors* are indicated by the subscripts $r{:}s$, which is the *index range*. This will result in a vector of size $s - r + 1$.

$$a_{r:s} = (a_r, \dots, a_s)$$

For example, if $a = (2, 4, 1, 8, 3)$, then

$$a_{2:4} = (4, 1, 8).$$

Note that the length of the above vector is $4 - 2 + 1 = 3$.

- In *vector addition* we sum corresponding entries. For example,

$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}.$$

- In *scalar-vector multiplication*, every element of the vector is multiplied by the scalar.

$$\beta a = (\beta a_1, \dots, \beta a_n)$$

For example,

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}.$$

- The *inner product* of two vectors is given by

$$a^T b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + \cdots + a_n b_n.$$

For example,

$$(-1, 2, 2)^T (1, 0, -3) = -1 + 0 - 6 = -7.$$

Flops

Basic arithmetic operations ($+$, $-$, $*$, $/$) are called **floating point operations** or **flops**. A crude approximation of a program runtime is given by

$$\text{runtime} \approx \frac{\text{number of operations (flops)}}{\text{computer speed (flops per second)}}.$$

The **operation count** (**flop count**) is the total number of operations in an algorithm. The dominant term is the highest-order term in the flop count. For example, an inner product takes $2n - 1$ flops. The dominant term is $2n$ so the algorithm runtime is of order n (we ignore the constants).

Linear functions

Lecture recap

- A *linear* function $f : \mathbf{R}^n \to \mathbf{R}$ is just an inner product, and can be expressed as

$$f(x) = a^T x.$$

- An *affine* function $f : \mathbf{R}^n \to \mathbf{R}$ is a linear function with an offset, and can be expressed as

$$f(x) = a^T x + b.$$

- *Superposition property* of linear and affine functions.
  - If f is linear
  $$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y).$$
  - If f is affine
  $$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \text{ if } \alpha + \beta = 1.$$

- Relationship between affine and linear functions:

All linear functions are affine, but not all affine functions are linear.

- The (first-order) Taylor approximation of a function $f(x)$ at a point z is given by

$$\hat{f}(x) = f(z) + \nabla f(z)^T (x - z).$$

Norm and distance

Lecture Recap

- The *norm* of a vector is the square root of the inner product of a vector with itself,

$$\|x\| = \sqrt{x^T x}.$$

- Main properties:
  - *Nonnegative homogeneity*: $\|\beta x\| = |\beta| \|x\|$.
  - *Triangle inequality*: $\|x + y\| \le \|x\| + \|y\|$.
  - *Nonnegativity*: $\|x\| \ge 0$.
  - *Definiteness*: $\|x\| = 0 \iff x = 0$.
  - *Cauchy-Schwarz inequality*: $|a^T b| \le \|a\| \|b\|$.

- Distance:
  - $\mathbf{dist}(a, b) = \|a - b\| = \|b - a\|$.
  - Distance squared:

$$\begin{aligned} \|a - b\|^2 &= (a - b)^T (a - b) \\ &= a^T (a - b) - b^T (a - b) \\ &= a^T a - a^T b - b^T a + b^T b \qquad (\text{distributive: } (a + b)^T c = a^T c + b^T c) \\ &= \|a\|^2 - 2a^T b + \|b\|^2 \qquad (\text{commutative: } a^T b = b^T a) \end{aligned}$$

- The *root-mean-square* of an n-vector x is given by

$$\mathbf{rms}(x) = \frac{\|x\|}{\sqrt{n}}.$$

- The *standard-deviation* of an n-vector x is given by

$$\mathbf{std}(x) = \frac{\|x - \mathbf{avg}(x)\mathbf{1}\|}{\sqrt{n}} = \frac{\left\| x - \frac{\mathbf{1}^T x}{n} \mathbf{1} \right\|}{\sqrt{n}}.$$

- The *angle* between two vectors a and b is given by

$$\angle(a, b) = \arccos\left(\frac{a^T b}{\|a\| \|b\|} \right).$$

Clustering notation

The k-means clustering algorithm is given below:

> Given $x_1, \dots, x_N \in \mathbf{R}^n$ and $z_1, \dots, z_k \in \mathbf{R}^n$
> **repeat**
> *Update partition:* assign i to G_j, $j = \arg\min_{j'=1,\dots,k} \|x_i - z_{j'}\|^2$
> *Update centroids:* $z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$
> **until**
> $z_1, \dots, z_k$ stop changing

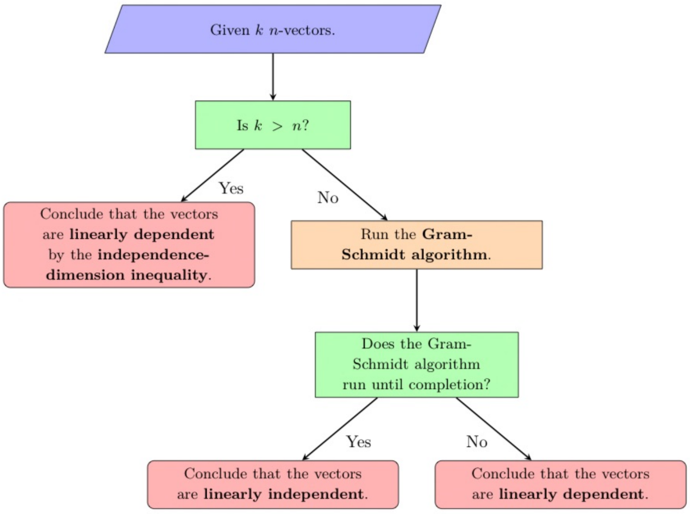
Linear independence

Lecture recap

- A set of k n-vectors is said to be *linearly dependent* if $\beta_1 a_1 + \cdots + \beta_k a_k = 0$ holds for some $\beta_1, \dots, \beta_k$ that are not all 0.
- A set of k n-vectors is said to be *linearly independent* if $\beta_1 a_1 + \cdots + \beta_k a_k = 0$ only when $\beta_1 = \cdots = \beta_k = 0$.
- A subset of a set of linearly independent vectors is also linearly independent.
- A superset of a set of linearly dependent vectors is linearly dependent.
- An orthonormal set of vectors satisfies the following:

$$a_i^T a_j = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

- A set of k n-vectors is said to form a *basis* if the following conditions are satisfied:
  - There are $k = n$ n-vectors.
  - They are linearly independent.
- How to determine if a set of vectors is linearly independent.



Matrices

Lecture recap

- A $m \times n$ matrix is a rectangular array of numbers containing m rows and n columns.
  - We say that $A \in \mathbf{R}^{m \times n}$.
  - The entry in the ith row and jth column is denoted A_{ij}.
- A matrix is said to be:
  - *tall* when $m > n$.
  - *square* when $m = n$.
  - *wide* when $m < n$.
- Some special matrices:
  - *Identity matrix*, denoted I_n is square and contains 1s on the diagonal and 0s elsewhere.
  - *Zero matrix* $0_{m \times n}$.
  - *Diagonal matrix* if $A_{ij} = 0$ for $i \neq j$.
  - *Triangular matrices* are square, and there are two types:
    - Lower triangular if $A_{ij} = 0$ for $i < j$.
    - Upper triangular if $A_{ij} = 0$ for $i > j$.
- Operations:
  - Block matrices (every block row has the same number of rows, and every block column has the same number of columns).
  - Transposition: $A_{ij} = (A^T)_{ji}$.
  - Matrix addition:
    - Commutative: $A + B = B + A$.
    - Associative: $(A + B) + C = A + (B + C)$.
    - Transpose of sum: $(A + B)^T = A^T + B^T$.
  - Scalar matrix multiplication:
    - Distributive: $(\beta + \gamma)A = \beta A + \gamma A$.
    - "Associativity" of scalar multiplication: $(\beta\gamma)A = \beta(\gamma A)$.
  - (Frobenius) norm of matrix: $\|A\| = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$.
  - *Matrix vector multiplication*: $Ax = y$.
    - $A \in \mathbf{R}^{m \times n}$, $x \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.
    - Row interpretation: $y_i = b_i^T x$ when b_i^T is the ith row of A.
    - Column interpretation: $y = \sum_{j=1}^{n} x_j a_j$ where a_j is the jth column of A.