1 Approximate Retrieval

Dataset $S \subset \mathbb{R}^D$, distance function $d : \mathbb{R}^D \times \mathbb{R}^D \leftarrow \mathbb{R}$

Nearest neighbour $s^* = \arg\min_{s \in S} d(q, s)$ for query q

Near duplicate detection: find all s,s' in S with distance at most ϵ

1.1 Distance Function $d : S \times S \to \mathbb{R}$

$\forall s, t \in S : d(s, t) \geq 0$

$\forall s : d(s, s) = 0$

$\forall s, t : d(s, t) = d(t, s)$

$\forall s, t, r : d(s, t) + d(t, r) \geq d(s, r)$

Cosine Distance: $d(x, y) = arccos \frac{x^T y}{||x||_2 ||y||_2}$

2 Linear Algebra

2.1 Vector Norms

are positive scalable, full-fill the triangular inequality, norm of 0 is 0

2.1.1 p-Norm

$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$

2.1.2 Euclidean Norm

p-Norm where $p = 2$

2.1.3 1-Norm

Manhattan-Norm $||x||_1 = \sum_{i=1}^{n} |x_i|$

2.1.4 Zero-Norm

counts the number of non-zero entries.

2.2 Matrix Norms

2.2.1 Nuclear Norm

$||.||_*$ sum of singular values

2.2.2 Frobenious-Norm

$sqrt(sum(sum(A.^2)))$

2.2.3 Spectral Norm

Largest singular value if square

$||A||_2 = \sigma_{max}(A)$ Is equals to the 2-Norm

2.2.4 Induced Matrix Norms

$||A|| = max\left(\frac{||Ax||}{||x||}\right)$

2.3 Orthogonality

2.3.1 Vectors

inner (scalar) product $\langle \, . \, , \, . \, \rangle = 0$

2.3.2 Matrices

quadratic, values are in $\mathbb{R}$, $Q^T = Q^{-1}$

2.3.3 Functions

$f(x)$ orth. to $g(x)$ if $0 = \int f(x)g(x)dx$

2.3.4 Coherence

$m(U) = max_{i,j:i \neq j} |u_i^T u_j|$

2.3.5 Convexity

$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$