HDP
(14) Johnson-Lindenstrauss Lemma

@ 滕佳烨

前情回顾:

$f(X) \to \mathbb{E}f(X)$ with f Lipschitz and $X \sim \mathbb{P}$

- sphere with Euclidean metric
- Gaussian RV with Euclidean metric
- Bernoulli RV with Hamming metric

→ Johnson-Lindenstrauss Lemma

Johnson-Lindenstrauss Lemma

Theorem 5.3.1 (Johnson-Lindenstrauss Lemma). *Let $\mathcal{X}$ be a set of N points in $\mathbb{R}^n$ and $\varepsilon > 0$. Assume that*

$$m \geq (C/\varepsilon^2) \log N.$$

Consider a random m-dimensional subspace E in $\mathbb{R}^n$ uniformly distributed in $G_{n,m}$. Denote the orthogonal projection onto E by P. Then, with probability at least $1 - 2\exp(-c\varepsilon^2 m)$, the scaled projection

$$Q := \sqrt{\frac{n}{m}} P$$

is an approximate isometry on $\mathcal{X}$:

$$(1-\varepsilon)\|x-y\|_2 \leq \|Qx - Qy\|_2 \leq (1+\varepsilon)\|x-y\|_2 \quad \text{for all } x, y \in \mathcal{X}. \tag{5.10}$$

Johnson-Lindenstrauss Lemma

\# Points = N
Orthogonal projection from Dim n [high] to Dim m [low]
Dimension: $m = \Omega(\log N)$ [not too low]
Then w.h.p, the projection Q is approximate isometry
[the original data structure does not change too much]

$$(1-\varepsilon)\|x-y\|_2 \leq \|Qx-Qy\|_2 \leq (1+\varepsilon)\|x-y\|_2 \quad \textit{for all } x, y \in \mathcal{X}.$$

Proof Hint: Transfer the randomness on Q to (x-y), then x-y is uniform on the sphere. Directly apply concentration of f(x) where $f(x) = ||Px||_2$

Take-away Messages

JL lemma: when we do orthogonal projection from high dim to low dim ($\Omega(\log N)$), the distance between two data points does not change much.

@ 滕佳烨

Thanks!