Dimensionality Reduction in Euclidean Norm

Piotr Indyk

Recap

- Have seen several algorithms for Hamming metric $\{0,1\}^d$
- All of them obtained by dimensionality reduction
- Time to proceed to "richer" norms:
  - Euclidean norm $\|\cdot\|_2$ (denoted by l_2^d)
  - Manhattan norm $\|\cdot\|_1$ (denoted by l_1^d):

$$\|x-y\|_1 = \sum_{i=1}^{d} |x_i - y_i|$$

Extension to l_1

Distance-preserving mapping $l_1^d \rightarrow$ Hamming cube:

- round the coordinates to $\{1 \dots M\}$
- embed $\{1 \dots M\}^d$ into $\{0,1\}^{dM}$ via unary mapping:

$$
\begin{array}{lcl}
(x_1, \dots, x_d) & \rightarrow & \underbrace{1 \dots\dots 1}_{x_1}\underbrace{0 \dots 0}_{M-x_1} \dots \underbrace{1 \dots 1}_{x_d}\underbrace{0 \dots\dots 0}_{M-x_d} \\
(y_1, \dots, y_d) & \rightarrow & \underbrace{1 \dots 1}_{y_1}\underbrace{0 \dots\dots 0}_{M-y_1} \dots \underbrace{1 \dots 1}_{y_d}\underbrace{0 \dots\dots 0}_{M-y_d}
\end{array}
$$

- Huge dimension, but can reduce it later

What about Euclidean norm ?

"Big picture"

- We show a randomized mapping $A : \Re^d \rightarrow \Re^k$, such that for any x, we will be able to estimate $\|x\|_2$ from $A(x)$

- The mapping A is linear. Therefore, given Ap and Aq, we can compute $Ap - Aq = A(p - q)$; by setting $x = p - q$ and applying the above, we can estimate $\|p - q\|_2$ from Ap and Aq.

- Note that it is not necessarily the case that we estimate $\|x\|_2$ by computing $\|Ax\|_2$. In fact, more complex estimators will be used.

- We will show how to extend the approach to other norms (e.g., l_1).

Normal distribution

Gaussian (or normal) distribution $N(0,1)$:

- Range: $(-\infty, \infty)$
- Density: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- Mean$= 0$, Variance$= 1$
  (in general these are the parameters)
- Key fact: X, Y independent Gaussians
  $\Rightarrow Z = X + Y$ also a Gaussian

Basic probability facts:

- $\text{Var}(cX) = c^2 \text{Var}(X)$
- X, Y independent
  $\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

“The Gaussian Method”

- Take $x \in \Re^d$
- Choose a vector $a = (a_1, \ldots, a_d)$, a_i's are independently drawn from $N(0,1)$
- Consider $s = a \cdot x = \sum_{i=1}^{d} a_i x_i$
  - Each term $a_i x_i$ is a Gaussian with variance x_i^2
  - $s = \sum_{i=1}^{d} a_i x_i$ is a Gaussian with variance $\sum_i x_i^2$
  - Thus, s is distributed as $(\sum_i x_i^2)^{1/2} Z = \|x\|_2 Z$, where $Z \in N(0,1)$
- If we take a $k \times d$ matrix $A = [a^1 a^2 \ldots a^k]^T$, then Ax consists of k independent samples, each distributed like $\|x\|_2 Z$, $Z \in N(0,1)$
- Can estimate the factor from the samples

Estimation

- Given: a set of samples $s_1, \ldots s_k$, taken independently from the same distribution as $c \cdot Z, Z \in N(0,1)$
- Goal: estimate c

Median estimator:

- A median of a random variable X: a value m such that

$$Pr[X > m] = Pr[X < m] = 1/2$$

- Let m be the median of $|Z|$, $Z \in N(0,1)$
- The median of $|cZ|$ is equal to cm
- Suffices to estimate the median

Median estimator

Fact: Let X be any* r.v. taking real values.
For $k = C \log(1/\delta)/\epsilon^2$ (C large constant) let $s_1 \dots s_k$ be independently chosen from the same distribution as X.
Then $M = median(s_1, \dots, s_k)$ satisfies

$$1/2 - \epsilon \leq Pr[X < M] \leq 1/2 + \epsilon$$

with probability $\geq 1 - \delta$.

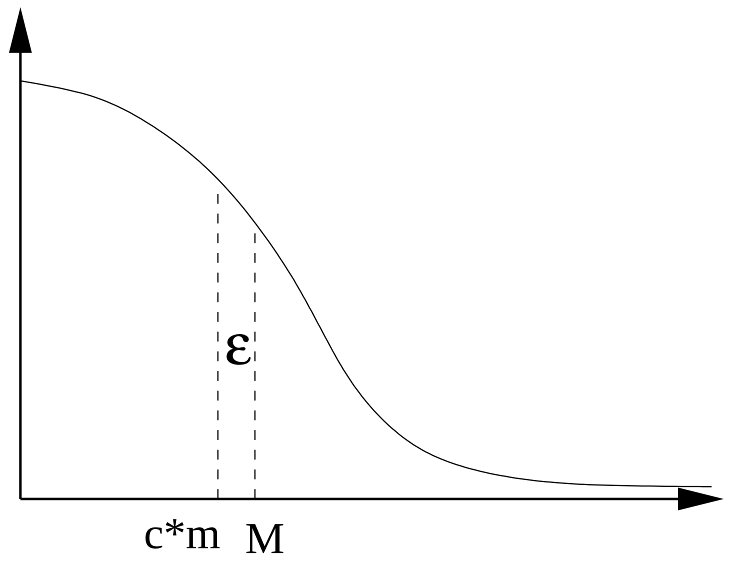
Proof: Use Chernoff bound.

Estimating c

- Median estimator gives us M, such that

$$1/2 - \epsilon \leq \Pr[|cZ| < M] \leq 1/2 + \epsilon$$

- Density of $|cZ|$ looks like this:



- Thus, for certain constant D

$$cm(1 - D\epsilon) \leq M \leq cm(1 + D\epsilon)$$

Final Theorem

Theorem: There is a constant B such that for any $\epsilon, \delta > 0$ and any $x \in \Re^d$, if we take a $k \times d$ matrix A with entries drawn independently from $N(0,1)$ where $k = B \log(1/\delta)/\epsilon^2$, then

$$median(|Ax|)/m = \|x\|_2(1 \pm \epsilon)$$

with probability at least $1 - \delta$.

Corollary: There is a randomized mapping $A : \Re^d \to \Re^k$ such that for any $p, q \in \Re^d$, we can recover $\|p - q\|_2$ from Ap and Aq, up to a factor of $(1 - \epsilon)$.

Proof: $Ap - Aq = A(p - q) = Ax$, and we can estimate $\|x\|_2$.

Comments

- Positives:
  - Can compute short sketches for the l_2 norm.
  - Advance knowledge of r not required (unlike for Hamming cube)
  - Thus, we can get a $(1+\epsilon)$-approximate algorithm for the nearest neighbor problem in l_2^d with $O((n+d)\log n/\epsilon^2)$ query time:
    * Compute Ap for all data points p
    * To answer a query q, find p minimizing $median(|Ap - Aq|)$

- Negatives:
  - Use $median(|Ax|)$ to estimate the norm, instead of $\|Ax\|_2$
  - Works *only* for the l_2 norm

Norm Estimator

Theorem (Johnson-Lindenstrauss): $\|x\|_2$ can be approximated by $\|Ax\|_2/k$, with the same parameters. In particular, $k = O(\log(1/\delta)/\epsilon^2)$ suffices.

Proof intuition:

- Compute the expectation of $\|Ax\|^2 = \sum_{j=1}^{k} |a^j \cdot x|^2$:

$$\begin{aligned} E[\|Ax\|^2] &= \sum_{j=1}^{k} E[|a^j \cdot x|^2] \\ &= kE[|a \cdot x|^2] \\ &= k\|x\|^2 E[|Z|^2] = k\|x\|^2 \end{aligned}$$

  where $Z \in N(0,1)$. Note that $E[|Z|]^2$ is just the variance of Z, i.e., $= 1$.

- By Chernoff-like bounds, $\|Ax\|^2$ is likely to be close to the expectation.

Stable Distributions

Definition: A distribution D is p-*stable*, if for any $a, b \in \Re$ and $X, Y \in \mathcal{D}$, $aX + bY$ has same distribution as $(|a|^p + |b|^p)^{1/p}Z$, $Z \in \mathcal{D}$.

Fact: For any $p \in (0, 2]$, p-stable distribution exists (and random variables from this distribution can be generated).

Corollary: The median estimator can be used to estimate $\|x\|_p$, $p \in (0, 2]$.

References

- Median estimator for l_p norms: P. Indyk, "Stable Distributions, Pseudorandom Generators, Embeddings and Data Stream Computation", FOCS 2000. Available at my web page.

- Norm estimator for l_2, i.e., Johnson-Lindenstrauss lemma:
  - Original proof: W.B. Johnson and J. Lindenstrauss, "Extensions of Lipshitz mapping into Hilbert space", Contemporary Mathematics (26), 1984, p. 189-206.
  - Simplification of the original proof: P. Frankl and H. Maehara, "The Johnson-Lindenstrauss Lemma and the Sphericity of Some Graphs", Journal of Combinatorial Theory B (44), 1988, p. 355-362.
  - Gaussian-based variant of the proof: P. Indyk and R. Motwani, "Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality", 30th Symposium on Theory of Computing , 1998 (appendix). Available at my web page.