Lecture 33 Johnson-Lindenstrauss Lemma.

Motivation In this digital era, lots of data are transmited as we speak. Many data (such as images, videos) can be represented by high-dim. vectors. To speed up computation, it is of great practical importance to try to reduce the dimension.

Application: • clustering

• regression analysis. ....

The basic task we want to do is to tell distinct vectors apart. Use pts in $\mathbb{R}^d$ to represent data and the euclidean distance between pts measures their similarity.

$x \in \mathbb{R}^d$

Write $\|x\| := \|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$. l_2-norm.

euclidean distance btw. x & y : $\|x-y\|$.

Lem (Johnson-Lindenstrauss)

Given $0 < \varepsilon < 1$ and a set X of n pts in $\mathbb{R}^d$, then there exist a linear map $f: \mathbb{R}^d \to \mathbb{R}^m$, where $m = O\left(\frac{\log n}{\varepsilon^2}\right)$ s.t. $\forall\, u, v \in X$,

$$\|f(u) - f(v)\| = \underbrace{(1 \pm \varepsilon)\,\|u-v\|}$$

$$(1-\varepsilon)\|u-v\| \le \|f(u)-f(v)\| \le (1+\varepsilon)\|u-v\|$$

$$f = \begin{pmatrix} & & \\ & & \end{pmatrix}_{m\times d} \begin{pmatrix} u \\ \\ \end{pmatrix}_{d\times 1} \longrightarrow f(u) \in \mathbb{R}^m$$

Idea: Project pts randomly to a low dim. subspace.

Naive: $\mathbb{R}^d \longrightarrow \mathbb{R}^m$ $\quad$ unif random m coord. out of d $\quad (\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot)$

Ex: $u = (1, 0, \cdots, 0)$
$v = (0, 1, 0, \cdots, 0)$

to preserve dist., these m random coord. need to include 1^{st} or 2^{nd} coord., which is quite unlikely if $d \gg m$.

Lem (Distribution JL lemma)

Given $0 < \varepsilon < 1$, there exist C s.t. the following holds.

Let A be an $m \times d$ random matrix, in which each entry is a normal r.v. $\sim \frac{1}{\sqrt{m}} \cdot N(0,1)$ indep. of others, where $m \le C \cdot \varepsilon^{-2} \cdot \log\frac{1}{\delta}$. Then for any vector $x \in \mathbb{R}^d$,

$$\mathbb{P}\left(\|\overset{\mathbb{R}^m}{A x}\| = (1 \pm \varepsilon)\|\overset{\mathbb{R}^d}{x}\| \right) \ge 1 - \delta$$

To get the original JL lemma, we apply the distributional JL to all $\binom{n}{2}$ pairwise distance in X: let $A = f$, $f(u) - f(v) = Au - Av$

$$= A(u-v)$$

Choose $p \geq k$ $\delta = \frac{\delta'}{\binom{n}{2}}$ δ' small.

$$m = O\left(\varepsilon^{-2} \log \frac{1}{\delta}\right) = O\left(\varepsilon^{-2} \log n\right).$$

Pf: Fix $x \in \mathbb{R}^d$. We want to show that whp length is preserved: $\|Ax\| \approx \|x\|$

First we show that

Claim: $\mathbb{E}\|Ax\|^2 = \|x\|^2$.

$$\|Ax\|^2 = \sum_{i=1}^{m} \langle g^{(i)}, x\rangle^2$$

$$= m \cdot \langle g, x\rangle^2$$

$g \in \mathbb{R}^d$

$g = (g_1, \dots, g_d)$

Pf: $A = \begin{pmatrix} — g^{(1)} — \\ \vdots \\ — g^{(m)} — \end{pmatrix}_{m \times d}$

Let $g \in \mathbb{R}^d$, $g = (g_1, \dots, g_d)$, where $\boxed{g_i \sim \frac{1}{\sqrt{m}} N(0,1)}$ $\sim N(0, \frac{1}{m})$

$$\Rightarrow \|Ax\|^2 = m \cdot \langle g, x\rangle^2$$

$$\langle g, x\rangle = \sum_{i=1}^{d} g_i x_i$$

$$\mathbb{E}\langle g, x\rangle = \sum_{i=1}^{d} x_i \, \mathbb{E} g_i = 0$$

$$\Rightarrow \mathbb{E}\langle g, x\rangle^2 = \mathrm{Var}\langle g, x\rangle = \sum_{i=1}^{d} x_i^2 \, \mathrm{Var}[g_i]$$
$$= \frac{1}{m}\|x\|^2$$

$$\mathbb{E}\|Ax\|^2 = \mathbb{E}\left(m \, \langle g, x\rangle^2\right) = \|x\|^2 .$$

In particular, $\|Ax\|^2 = \sum_{i=1}^{m} \langle g^{(i)}, x\rangle^2$, where

$$\langle g^{(i)}, x\rangle \sim \frac{1}{\sqrt{m}}\|x\| \cdot \mathcal{N}(0,1)$$

chi-squared distribution

So $\|Ax\|^2$ is a χ^2 random variable w./ m degree of freedom. For such r.v.

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq \varepsilon \, \mathbb{E}X\right) \leq 2e^{-\varepsilon^2 m/8} \leq \delta$$

Set $m = O\left(\log(1/\delta)/\varepsilon^2\right)$ $\blacksquare$

Application to regression analysis.

Setup: Given data pts

$a_1, a_2, \ldots, a_n \in \mathbb{R}^d \longrightarrow$  $\xrightarrow[\text{response}]{\text{outcome}}$ $y_1, y_2, \ldots, y_n \in \mathbb{R}$

Want to figure out the relation between the data & the outcome.

$$\begin{pmatrix} \text{——} & a_1 & \text{——} \\ \text{——} & a_2 & \text{——} \\ & \vdots & \\ \text{——} & a_n & \text{——} \end{pmatrix} = A_{n \times d} \qquad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$$

Whether $\exists\, x \in \mathbb{R}^d$ s.t. $A \cdot x \approx y$

Least square regression

Goal:

$$\min_{x \in \mathbb{R}^d} \| Ax - y \|^2$$

Idea: Subspace embedding & union bound over ε-net.