Johnson-Lindenstrass Transform （JL 变换）

秦睿哲

【JL-定理】令 $\varepsilon \in (0,1)$，给定 R^d 中的包含 n 个点的点集 P，$k = \frac{1}{\varepsilon^2}\log n$（与 d 无关），那么存在一个从 R^d 到 R^k 的 Lipshcitz 映射 f，对于 $\forall u, v \in P$ 有

$$(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2$$

Lipshcitz 映射 f：构造一个矩阵 $B \in R^{k\times d}$，得到 $f(u) = Bu$。

映射的构造方法：$B = \frac{1}{\sqrt{k}}A$，其中 A 是一个 $k\times d$ 维矩阵，且它的每一个元素均独立同分布于高斯分布 $N(0,1)$。

【JL-定理的证明】

我们的证明过程主要参考了（[1] Random Projections）的思想。

要证明对于 $\forall u, v \in P$ 有 $(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2$，将 $u-v$ 看作 x，也就是要证明 $\left\|\frac{1}{\sqrt{k}}A\cdot x\right\|^2 \in (1\pm\varepsilon)\|x\|^2$。令 $y = \frac{1}{\sqrt{k}}Ax$，首先证明 $E[\|y\|^2] = \|x\|^2$。

记 $A = \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix}$，其中 A_j 是 A 的第 j 行，则 y 的第 j 个元素是 $\frac{1}{\sqrt{k}}A_j x$。

$$\|y\|^2 = \frac{1}{k}\sum_{j=1}^{k}(A_j\cdot x)^2$$

$$\Rightarrow E[\|y\|^2] = \frac{1}{k}\sum_{j=1}^{k}E[(A_j x)^2]$$

$$\begin{aligned} E[(A_j x)^2] &= E[\left(\sum_{i=1}^{d}A_{ji}x_i\right)^2] \\ &= E[\sum_{i,i'}A_{ji}A_{ji'}x_i x_{i'}] \end{aligned}$$

注意到当$i \neq i'$时$E[A_{ji}A_{ji'}x_i x_{i'}]=0$，因此

$$E[(A_j x)^2]=E[\sum_{\mathrm{i=1}}^{\mathrm{d}} A_{ij}^2 x_i^2]$$
$$=\sum_{i=1}^{d} x_i^2=\|x\|^2$$

所以$E[\|y\|^2]=\|x\|^2$.

接下来，目标是证明$\mathrm{Prob}\left(\|\mathrm{y}\|^2 \geq (1+\varepsilon)\|x\|^2\right)$足够小。

$$\|y\|^2=\frac{1}{k}\|Ax\|^2 \geq (1+\varepsilon)\|x\|^2 \quad \Leftrightarrow \quad \frac{\|Ax\|^2}{\|x\|^2} \geq (1+\varepsilon)k$$

$\frac{\|Ax\|^2}{\|x\|^2}=\sum_{j=1}^{k}\frac{(A_j x)^2}{\|x\|^2}=\sum_{j=1}^{k} Z_j^2$，这里我们定义$Z_j=\frac{A_j x}{\|x\|}$。

因此，我们的目标等价于证明$\mathrm{Prob}\left(\sum_{\mathrm{j=1}}^{\mathrm{k}} Z_j^2 \geq (1+\varepsilon)k\right)$足够小。

$$\mathrm{Prob}\left(\sum_{\mathrm{j=1}}^{\mathrm{k}} Z_j^2 \geq (1+\varepsilon)k\right)$$
$$=\mathrm{Prob}\left(e^{\lambda\sum_{\mathrm{j=1}}^{\mathrm{k}} Z_j^2} \geq e^{(1+\varepsilon)k\lambda}\right)$$
$$\leq \frac{1}{e^{(1+\varepsilon)k\lambda}} E[e^{\lambda\sum_{\mathrm{j=1}}^{\mathrm{k}} Z_j^2}]$$
$$\leq \frac{1}{e^{(1+\varepsilon)k\lambda}}\left(E[e^{\lambda Z_1^2}]\right)^k$$

*这里注意到$Z_j=\frac{A_j x}{\|x\|}=\sum_{i=1}^{d} A_{ji}\frac{x_i}{\|x\|}$符合高斯分布，因此

$$E[Z_j]=0, E[Z_j^2]=1. E[e^{\lambda Z_1^2}]=\frac{1}{\sqrt{1-2\lambda}}\text{。}$$

$$\Rightarrow \frac{1}{e^{(1+\varepsilon)k\lambda}}\left(E[e^{\lambda Z_z^2}]\right)^k$$

$$= \frac{1}{e^{(1+\varepsilon)k\lambda}}\left(\frac{1}{1-2\lambda}\right)^{\frac{k}{2}} \qquad (\lambda = \frac{\varepsilon}{2(1+\varepsilon)})$$

$$= \left((1+\varepsilon)e^{-\varepsilon}\right)^{\frac{k}{2}} \le e^{-\frac{k}{4}\left(\varepsilon^2-\varepsilon^3\right)}$$

所以得到 $\mathrm{Prob}\left(\|\mathrm{y}\|^2 \ge (1+\varepsilon)\|x\|^2\right) \le e^{-\frac{k}{4}\left(\varepsilon^2-\varepsilon^3\right)}$。

同理，$\mathrm{Prob}\left(\|\mathrm{y}\|^2 \le (1+\varepsilon)\|x\|^2\right) \le e^{-\frac{k}{4}\left(\varepsilon^2-\varepsilon^3\right)}$。

所以，$\mathrm{Prob}\left(\|\mathrm{y}\|^2 \in (1\pm\varepsilon)\|x\|^2\right) \ge 1-2e^{-\frac{k}{4}\left(\varepsilon^2-\varepsilon^3\right)}$。

【JL 变换与 PCA 的对比】

时间上，JL 为 $\Theta\left(n\cdot d\cdot\frac{\log n}{\varepsilon^2}\right)$ < PCA 的 $\Theta\left(n\cdot d^2\right)$。

JL 变换的优点为适合并行计算/分布式计算，通常用于流数据的操作和隐私保护等方面。

【其他 JL 变换】

• 基于高斯分布的 JL 变换的缺点：

投影矩阵为稠密矩阵，$A\cdot x$ 的计算时间为 $\Theta(k\cdot d)$。如果可以将 A 变成稀疏矩阵，那么计算时间与 A 中非零元素的个数相关。因此，为了优化 JL 变换，提出了其他的 JL 变换投影矩阵的构造方法。

• 其他 JL 变换的构造方法：

1、用 0/1 矩阵构造（[2] Database-friendly random projections: Johnson–Lindenstrauss with binary coins）

$$A \in R^{O\left(\frac{\log n}{\varepsilon^2}\right)\times d}$$

$$A_{ij}=\begin{cases}1 & \text{with prob } \frac{1}{2}\\ -1 & \text{with prob } \frac{1}{2}\end{cases} \quad 或者 \quad A_{ij}=\begin{cases}+1 & \text{with prob } \frac{1}{6}\\ 0 & \text{with prob } \frac{2}{3}\\ -1 & \text{with prob } \frac{1}{6}\end{cases}$$

*该方法实际效果比较好，运行时间较短。

2、Fast JL 变换 （[3] Faster dimension reduction）

投影矩阵 $\Phi = P\cdot H\cdot D,\ \ k=O\left(\frac{\log n}{\varepsilon^2}\right)$

其中(1) $P\in R^{k\times d}$ 是一稀疏矩阵 $P_{ij}=\begin{cases}N(0,\frac{1}{q}) & \text{with prob } q\\ 0 & \text{with prob } 1-q\end{cases}$

$$q=\min\left\{\Theta\left(\frac{\log^2 n}{d}\right),1\right\}\Rightarrow 与矩阵中的非零个数相关$$

(2) $H\in R^{d\times d}$ 是归一化的 Hadamard 矩阵

$H_{ij}=d^{-\frac{1}{2}}(-1)^{\langle i-1,j-1\rangle}$，其中 $\langle i,j\rangle$ 是用二进制表示的位向量 i,j 的点积（模 2）。

(3) $D\in R^{d\times d}$ 为一对角矩阵，每一个 D_{ii} 独立地从 $\{-1,1\}$ 中取值，概率均为 $\frac{1}{2}$。

参考文献：

1、Sham Kakade and Greg Shakhnarovich. Random Projections. CMSC 35900 (Spring 2009).
2、Achlioptas, D. Database-friendly random projections: Johnson–Lindenstrauss with binary coins. J. Comput. Syst. Sci. 66, 4 (2003), 671–687.
3、Ailon N, Chazelle B. Faster dimension reduction. Communications of the Acm, 2010, 53(2):97-104.