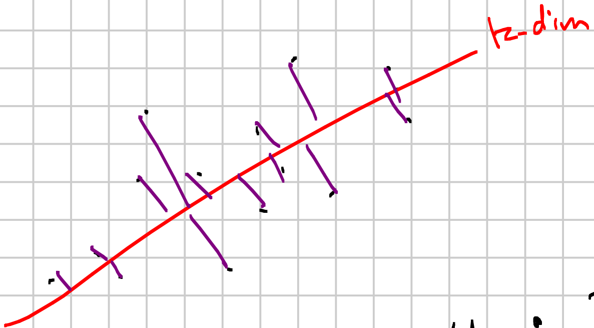
# Random Projections

Note Title 3/21/2016

<u>Input Data</u> point set $P \subset \mathbb{R}^d$




$d$ is very very large

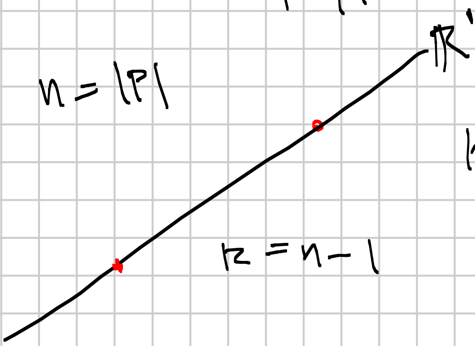
$\longrightarrow \mathbb{R}^k \qquad k << d$

500 100,000

$$\mu : \mathbb{R}^d \to \mathbb{R}^k$$

forall $p_1, p_2 \in P$ $\quad (1-\varepsilon)\|p_1 - p_2\| \le \|\mu(p_1) - \mu(p_2)\| \le \|p_1 - p_2\|(1+\varepsilon)$

$n = |P|$



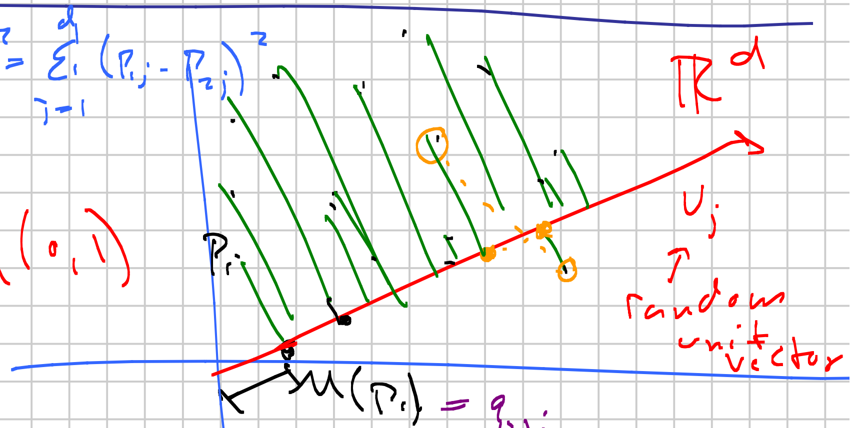

$k = n-1$

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

holds with probability $1-\delta$

---

$\mathbb{R}^d \to \mathbb{R}^{k=1}$ $\quad \|p_1 - p_2\|^2 = \sum_{j=1}^{d} (p_{1j} - p_{2j})^2$




$g_j = (g_{j1}, g_{j2}, \dots g_{jd}) \sim N_d(0,1)$

$g_{ji} \sim N(0,1)$

Box–Muller $(z_1, z_2)$

Uniform $[0,1]$

$u_j = \dfrac{g_j}{\|g_j\|}$

$q_{1j} = \langle u_j, p_1 \rangle$
$= \sum_{l=1}^{d} u_{jl} \cdot p_{1l}$

for $j = 1$ to $k$

Let $u_j \sim$ Random $d$-dim unit vector

Rescale : $u_j = \sqrt{\frac{d}{k}} \, \frac{u_j}{\|u_j\|}$

Define $\mu$

for each $p_i \in P$

for each $j$ in $1$ to $k$

$q_{ij} = \langle p_i , v_j \rangle$

$\mu(p_i) = q_i = (q_{i1}, q_{i2}, \dots, q_{ik}) \in \mathbb{R}^k$

Johnson-Lindenstrauss

$v_j \sim \{-1, 0, +1\}$

2/3