Lecture Notes 6

January 26, 2012

1 Low Distortion Embeddings

Suppose we have a set V of n randomly distributed points in $\mathbb{R}^d$. We want to find a mapping $f : \mathbb{R}^d \mapsto \mathbb{R}^k$ where $k \leq d$ and $k \sim \log n$ such that when we map two points $v, w \in V$ and map them into $\mathbb{R}^k$, then there is low distortion between $d(v, w)$ and $d(f(v), f(w))$. This function is described and its bound proved by the following theorem.

Theorem 1 (Johnson-Lindenstrauss)**.** *There exists a mapping $f : \mathbb{R}^d \mapsto \mathbb{R}^k$ such that for all $v, w \in V$*

$$(1-\epsilon)d(v,w) \leq d(f(v), f(w)) \leq (1+\epsilon)d(v,w)$$

with probability at most $1/n^3$.

Exercise: Prove that for $Pr(a\sigma \leq S_n - np \leq b\sigma) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, where $\sigma^2 = Var(S_n)$.

Proof. Consider two points in $v, w \in \mathbb{R}^d$ such that $d(v, w) = 1$. In order to translate the points into $\mathbb{R}^k$ we randomly the first k of its components, and discard the remaining $d - k$. If $v', w' \in \mathbb{R}^k$ are the mapping of v and w, then we will let $X = |v' - w'|$ be our random variable. Let $X = (x_1, \ldots, x^d)$ with the density function $f(x) = \frac{1}{2\pi^{d/2}} e^{-\sum_{i=1}^d x_i^2/2}$. Let $Y = (y_1, \ldots, y_d) = \frac{X}{||X||}$ be a random point on the unit sphere of $\mathbb{R}^d$. Let $Z = (y_1, \ldots, y_k)$ be the projection of Y into $\mathbb{R}^k$. We want to find and bound $L = ||Z||^2$.

Lemma 2. *E(L) = k/d.*

Proof.

$$E(L) = E(||Z||^2) = E\frac{(x_1^2 + \cdots + x_k^2)}{||X||^2} = \sum_{i=1}^k \frac{E(x_i^2)}{||X||^2} = \sum_{i=1}^k \frac{E(x_1^2)}{||X||^2} = \frac{kE(x_1^2)}{||X||^2}.$$

By linearity of expectation and since the x_i's are identically distributed. We know that

$$1 = ||Y||^2 = E(||Y||^2) = \frac{\sum_{i=1}^d E(x_i^2)}{||X||} = \frac{\sum_{i=1}^d E(x_1^2)}{||X||} = \frac{dE(x_1^2)}{||X||}$$

So $\frac{E(x_1^2)}{||X||} = 1/d$ which gives us $E(L) = k/d$. □

Claim 3.

$$E(e^{sx_1^2}) = \frac{1}{\sqrt{1-2s}}.$$

Proof. Since $X \sim \mathcal{N}(0,1)$, we have

$$E(e^{sx_1^2}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot e^{sx^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1-2s)x_1^2/2} dx = \frac{1}{\sqrt{2\pi(1-2s)}} \int_{-\infty}^{\infty} e^{-y^2/2} dx = \frac{1}{\sqrt{1-2s}}$$

when we substitute $y = x_1\sqrt{1-2s}$. □

Lemma 4.

1. *If $\beta < 1$, then $Pr(L \leq \beta k/d) \leq \beta^{k/2}\left(1 + \frac{1-(1-\beta)k}{d-k}\right)^{(d-k)/2} \leq \exp\left(\frac{k}{2}(1-\beta+\log\beta)\right)$.*

2. *If $\beta > 1$, then $Pr(L \geq \beta k/d) \leq \beta^{k/2}\left(1 + \frac{1-(1-\beta)k}{d-k}\right)^{(d-k)/2} \leq \exp\left(\frac{k}{2}(1-\beta+\log\beta)\right)$.*

Proof. We will only prove part (1) of this lemma since the argument for part (2) is almost identical.

$$
\begin{aligned}
Pr(L \leq \beta k/d) &= Pr(||Z||^2 \leq \beta k/d) \\
&= Pr\left(\frac{x_1^2 + \cdots + x_k^2}{x_1^2 + \cdots + x_d^2} \leq \beta k/d\right) \\
&= Pr\left(d(x_1^2 + \cdots + x_k^2) \leq \beta k(x_1^2 + \cdots + x_d^2)\right) \\
&= Pr\left(\beta k(x_1^2 + \cdots + x_d^2) - d(x_1^2 + \cdots + x_k^2) \geq 0\right) \\
&= Pr\left(e^{t(\beta k(x_1^2 + \cdots + x_d^2) - d(x_1^2 + \cdots + x_k^2))} \geq e^{t \cdot 0} = 1\right) \\
&\leq E\left(e^{t(\beta k(x_1^2 + \cdots + x_d^2) - d(x_1^2 + \cdots + x_k^2))}\right) && \text{(Markov's Inequality)} \\
&= E\left(e^{t\beta k x_1^2 (d-k)}\right) \cdot E\left(e^{t(\beta k - d) x_1^2 k}\right) \\
&= \left(\frac{1}{1 - 2t\beta k}\right)^{\frac{d-k}{2}} \left(\frac{1}{1 - 2t(\beta k - d)}\right)^{\frac{k}{2}}. && \text{(Claim 3)}
\end{aligned}
$$

We want to find t to minimize the term above. Using calculus, we find a minimum at $t_0 = \frac{1-\beta}{2\beta(d-k\beta)}$. We get the second bound of the inequality in the lemma by using the fact that $1 + x \leq e^x$. □

If we set $k \geq \frac{3 \log n}{2\epsilon^2}$ and $\beta = 1 - \epsilon$, by Lemma 4 we get

$$
Pr(L \leq (1-\epsilon)k/d) \leq \exp\left(\frac{k}{2}(1 - \beta + \log \beta)\right) \leq e^{-k\epsilon^2/2} \leq \frac{1}{n^3}.
$$

□