Randomized Algorithms Georgia Tech CS8113F, Winter 1999
Leonard Schulman CCB 234, (404) 894-6438, schulman@cc.gatech.edu
Lecture 5, Jan 25 '99, "Johnson-Lindenstrauss Embedding (cont.)"
Scribe: Russell Martin

Recall our goal is to embed a set of n points from $\mathcal{R}^n$ into a lower dimensional space without distorting the distance between the points "too much."

We defined W to be a $k \times n$ matrix with rows $r_1, r_2, \ldots, r_k$ chosen independently from the spherically symmetric Gaussian density in $\mathcal{R}^n$. Then for $v \in \mathcal{R}^n$, $r_i \cdot v$ is normally distributed with standard deviation (proportional to) $||v||$. We write

$$||Wv||^2 = \sum_{i=1}^{k} (r_i \cdot v)^2$$

where the $(r_i \cdot v)^2$ are independent random variables with expected value $E((r_i \cdot v)^2) = ||v||^2$. (Equality comes from suitable choice of the scale of the spherically symmetric Gaussian.) Now we're ready to use the Chernoff bound.

1 Using the Chernoff Bound

We'll simplify by assuming that $||v|| = 1$. Moreover, since we can without loss of generality rotate v to the vector $(||v||, 0, ..., 0)$, we will think of only the first coordinate of the vector r (and abusing notation, we will continue to refer to this coordinate by r).

We now have independent random variables $r_1, r_2, \ldots, r_k$, each normally distributed with $E(r_i^2) = 1$ (so $E(\sum_{i=1}^k r_i^2) = k$). We want a deviation bound on $\sum_{i=1}^k r_i^2$.

Set random variables $y_i = r_i^2 - 1$ so that $E(y_i) = 0$. With this change of variables we now want a deviation bound on $\overline{y} = \frac{1}{k}\sum_{i=1}^k y_i$. We will use the Chernoff bound to help us.

To get a Chernoff bound, we need the moment generating function, $g(t)$, for y_i.

Recall that

$$P(\overline{y} > \varepsilon) < [\inf_{t>0} e^{-\varepsilon t} g(t)]^k \qquad \text{for } \varepsilon > 0, \tag{1}$$

and the same proof shows

$$P(\overline{y} < \varepsilon) < [\inf_{t<0} e^{-\varepsilon t} g(t)]^k \qquad \text{for } \varepsilon < 0. \tag{2}$$

We will need both bounds to show that the distances aren't distorted too much in either way, i.e. either too small or too large. So what is the moment generating function $g(t)$?

$$\begin{aligned} g(t) &= E(e^{ty}) \;=\; E(e^{t(r^2-1)}) \\ &= \int_{-\infty}^{\infty} e^{t(r^2-1)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2} dr \\ &= e^{-t} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2(1-2t)} dr \end{aligned}$$

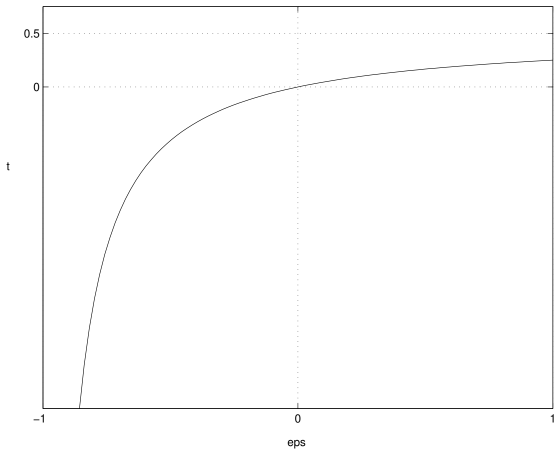
$$
\begin{aligned}
&= \frac{e^{-t}}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2t}}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2(1-2t)} dr \\
&= \frac{e^{-t}}{\sqrt{1-2t}}
\end{aligned}
$$

The last equality follows as the integrand is the density of a normal random variable with standard deviation $\frac{1}{\sqrt{1-2t}}$.

Thus, $g(t)$ is well defined and differentiable in $(-\infty, \frac{1}{2})$, with $g(0) = 1$ and $g'(0) = 0$.

For a given ε what t should be used in the Chernoff bounds (equations (1) and (2))? After some calculus, we find that $t = \frac{\varepsilon}{2(1+\varepsilon)}$ is the appropriate value.

Below, we give a sketch of t versus ε.

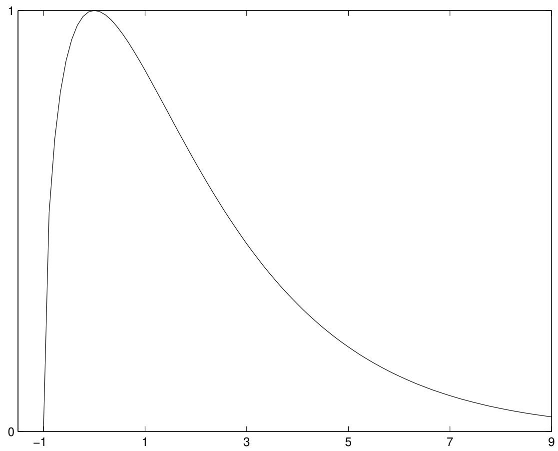


Plugging the value of t above into the Chernoff bound (1 and 2) we get

$$
P(\overline{y}/\varepsilon > 1) < ((1+\varepsilon)^{\frac{1}{2}} e^{-\frac{\varepsilon}{2}})^k = \exp((-\frac{1}{4}\varepsilon^2 + O(\varepsilon^3))k) \tag{3}
$$

(This holds for ε positive or negative.)

The function $(1+\varepsilon)^{\frac{1}{2}} e^{-\frac{\varepsilon}{2}}$ is shown in the sketch below.



Now let's apply this bound to the Johnson-Lindenstrauss embedding. We want everything up to $O(\varepsilon^2)$ terms.

For distortion $1+\varepsilon$ in the embedding we want to ensure that

$$||v||\frac{1}{\sqrt{1+\varepsilon}} \leq ||Wv|| \leq ||v||\sqrt{1+\varepsilon}$$

or

$$||v||^2\frac{1}{1+\varepsilon} \leq ||Wv||^2 \leq ||v||^2(1+\varepsilon).$$

From inequality (3) we have that, for $\varepsilon > 0$,

$$P(|\overline{y}| > \varepsilon) < 2e^{-[\frac{1}{4}\varepsilon^2+O(\varepsilon^3)]k}.$$

So finally we have

$$P(\frac{1}{\sqrt{1+\varepsilon}} \leq \frac{||Wv||}{||v||} \leq \sqrt{1+\varepsilon}) \leq 2\left(\begin{array}{c} n \\ 2 \end{array}\right)e^{-[\frac{1}{4}\varepsilon^2+O(\varepsilon^3)]k}.$$

This probability is strictly less than one provided $k = \frac{8\log n}{\varepsilon^2}(1+O(\varepsilon))$. To get a randomized "Las Vegas" algorithm simply by trying projections at random and testing whether their distortion is satisfactory, use say $k = \frac{8\log n}{\varepsilon^2}(1+O(\varepsilon))+1$ so that the probability of having distortion greater than $1+\varepsilon$ is at most $1/e$.

Note: Concerning another embedding question: Finite l_2 metric spaces can be embedded in l_1 with exact distance preservation — see lectures 7-8. (And there is an algorithm – deterministic, in fact – to find such an embedding, but it takes exponential time in the number of points in the space).