CS 4850 Mathematical Foundations for the Information Age	February 3, 2010
Lecture 5	
Lecturer: John Hopcroft	*Scribes: Arvind Kulkarni(avk34), Stanley Shen(ds676)*

1 Random projection theorem

We want to prove that we can do *dimension reduction.* Suppose we have data in $d = 25,000$ dimensions. For example, we have journal articles and we reference them by a column vector, and we try to cluster to do something. It would be convenient to reduce the dimension to $d = 50$ provided we do not lose much of the structures. It would be nice if for every pair of points the distance will be preserved except for a constant factor. Suppose we have n points, and we look at all n^2 distances, they will all shrink by the same amount. And for clusters, in lower dimensions we still get the same clusters as in higher dimensions.

Theorem 1 (Random Projection Theoremd) *Let z be a random unit length vector in d-dimensions and let $\tilde{z} = (z_1, z_2, \cdots, z_k)$. For $0 < \epsilon < 1$*

$$\mathbf{Pr}[||\tilde{z}|^2 - \frac{k}{d}| \geq \epsilon\frac{k}{d}] \leq \mathrm{e}^{-\frac{k\epsilon^2}{4}}$$

Proof Two cases:

Case 1 $|\tilde{z}|^2 \geq \frac{k}{d}$, $|\tilde{z}|^2 - \frac{k}{d} \geq \epsilon\frac{k}{d}$, $|\tilde{z}|^2 \geq (1+\epsilon)\frac{k}{d}$, $\beta \overset{\text{def}}{=} 1+\epsilon$, $|\tilde{z}|^2 \geq \beta\frac{k}{d}$

Case 2 $|\tilde{z}|^2 \leq \frac{k}{d}$, $\frac{k}{d} - |\tilde{z}|^2 \geq \epsilon\frac{k}{d}$, $|\tilde{z}|^2 \leq (1-\epsilon)\frac{k}{d}$ $\beta \overset{\text{def}}{=} 1-\epsilon$, $|\tilde{z}|^2 \leq \beta\frac{k}{d}$

We only prove Case 2. Case 1 can be proved in a similar way. Not that if x is normally distributed with mean zero and variance 1, we have

$$\begin{aligned}
\mathbf{E}[\mathrm{e}^{tx^2}] &= \int_{-\infty}^{\infty} \mathrm{e}^{tx^2} p(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{tx^2} \cdot \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-(\frac{1}{2}-t)x^2}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(1-2t)x^2}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{2\pi}\frac{1}{\sqrt{1-2t}} \qquad\qquad \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2a^2}}\,\mathrm{d}x = \sqrt{2\pi}a \\
&= \frac{1}{\sqrt{1-2t}}
\end{aligned}$$

Generate unit vector z, pick $x_1, x_2, \cdots, x_d$ using Gaussians. Let $z = \frac{x}{|x|}$. Caution: $x_1, \cdots, x_d$ are independent, but they lose independence after being normalized. Case 2: $|\tilde{z}|^2 < \frac{k}{d}$

$$
\begin{aligned}
\mathbf{Pr}[|\tilde{z}|^2 \le \beta\frac{k}{d}|z|^2] &= \mathbf{Pr}[|\tilde{z}|^2 \le \beta\frac{k}{d}] && |z| = 1 \\
&= \mathbf{Pr}[x_1^2 + x_2^2 + \cdots + x_k^2 \le \beta\frac{k}{d}(x_1^2 + x_2^2 + \cdots + x_d^2)] \\
&= \mathbf{Pr}[t(x_1^2 + x_2^2 + \cdots + x_k^2) \le \beta k(x_1^2 + x_2^2 + \cdots + x_d^2)] \\
&= \mathbf{Pr}[\beta k(x_1^2 + x_2^2 + \cdots + x_d^2) - d(x_1^2 + x_2^2 + \cdots + x_k^2) \ge 0] \\
&= \mathbf{Pr}[t(\beta k(x_1^2 + x_2^2 + \cdots + x_d^2) - d(x_1^2 + x_2^2 + \cdots + x_k^2)) \ge 0] && \forall t \ge 0 \\
&= \mathbf{Pr}[\mathrm{e}^{t(\beta k(x_1^2 + x_2^2 + \cdots + x_d^2) - d(x_1^2 + x_2^2 + \cdots + x_k^2))} \ge 1] \\
&\le \mathbf{E}[\mathrm{e}^{t(\beta k(x_1^2 + x_2^2 + \cdots + x_d^2) - d(x_1^2 + x_2^2 + \cdots + x_k^2))}] && \text{Markov's inequality} \\
&= \mathbf{E}[\mathrm{e}^{t(\beta k(x_{k+1}^2 + x_{k+2}^2 + \cdots + x_d^2) + \beta k(x_1^2 + x_2^2 + \cdots + x_k^2) - d(x_1^2 + x_2^2 + \cdots + x_k^2))}] \\
&= \mathbf{E}[\mathrm{e}^{t(\beta k(x_{k+1}^2 + x_{k+2}^2 + \cdots + x_d^2) + (\beta k - d)(x_1^2 + x_2^2 + \cdots + x_k^2))}] \\
&= \mathbf{E}[\mathrm{e}^{t\beta k(x_{k+1}^2 + x_{k+2}^2 + \cdots + x_d^2)}]\mathbf{E}[\mathrm{e}^{t(\beta k - d)(x_1^2 + x_2^2 + \cdots + x_k^2)}] && x_1, x_2, \cdots, x_d \text{ are independent} \\
&= [\mathbf{E}[\mathrm{e}^{t\beta k x_1^2}]]^{d-k}[\mathbf{E}[\mathrm{e}^{t(\beta k - d)x_1^2}]]^k \\
&= (\frac{1}{\sqrt{1 - 2t\beta k}})^{d-k}(\frac{1}{\sqrt{1 - 2t(\beta k - d)}})^k \\
&= (\frac{1}{1 - 2t\beta k})^{\frac{d-k}{2}}(\frac{1}{1 - 2t(\beta k - d)})^{\frac{k}{2}} \\
&\stackrel{\text{def}}{=} g(t)
\end{aligned}
$$

Becasue t is arbitrary positive real number. We can find the minimum of $g(t)$ to get a tighter upper bound. Minimizing $g(t)$ is the same as maximizing $f(t) \stackrel{\text{def}}{=} (1 - 2t\beta k)^{\frac{d-k}{2}}(1 - 2t(\beta k - d))^{\frac{k}{2}}$, which is equivalent to maximizing $h(t) \stackrel{\text{def}}{=} \ln f(t)$. Calculate the derivative $h'(t)$ and let $h'(t) = 0$. It is easy to get $t_0 = \frac{\beta - 1}{2\beta(\beta k - d)}$, plug it in, we have

$$
\begin{aligned}
\mathbf{Pr}[|\tilde{z}|^2 \le \beta\frac{k}{d}] &\le g(t_0) \\
&= \beta^{\frac{k}{2}}(\frac{d - \beta k}{d - k})^{\frac{d-k}{2}} \\
&= \beta^{\frac{k}{2}}(\frac{d - k + k\beta k}{d - k})^{\frac{d-k}{2}} \\
&= \beta^{\frac{k}{2}}(1 + \frac{k\beta k}{d - k})^{\frac{d-k}{2}} \\
&= \beta^{\frac{k}{2}}\mathrm{e}^{\frac{k(1-\beta)}{2}} \\
&= \mathrm{e}^{\frac{k}{2}\ln\beta + \frac{k}{2}(1-\beta)} \\
&= \mathrm{e}^{\frac{k}{2}(\ln\beta + 1 - \beta)} \\
&= \mathrm{e}^{\frac{k}{2}(\ln(1-\epsilon) + \epsilon)} \\
&\le \mathrm{e}^{-\frac{k}{4}\epsilon^2} && \ln(1-\epsilon) \le -\epsilon - \frac{1}{2}\epsilon^2 \ \forall \epsilon \in (0, 1)
\end{aligned}
$$

■

If we let $k = \frac{64 \ln n}{\epsilon^2}$, we have $\mathbf{Pr}[||\tilde{z}|^2 - \frac{k}{d}| \ge \epsilon\frac{k}{d}]$ is upper bounded by $\mathrm{e}^{-16\ln n} = n^{-16}$(in fact $\frac{1}{n^4}$ suffices). If we have n points, there are $\binom{n}{2} \approx n^2$ pairs. By the union bound[1], the probability that any pair is not

[1] $\Pr[\cup_{i\in S} A_i] \le \sum_{i\in S} \Pr[A_i]$.

preserved within $1+\epsilon$ factor is upper bounded by $\frac{1}{n}$. So we get the desired reduction *almost surely*[2]. We can use the random projection theorem to prove the Johnson-Lindenstrauss lemma. The sketch is the above.

[2] *Almost surely* means the the event happens with probability one as n approaches ∞