Feynman Liang
STAT 210B
Lecture #5
1/31/2017

1 Johnson-Lindenstrauss wrap up

Setup:

(a) $N \geq 2$ points in $\mathbb{R}^d$

(b) $\{u^1, u^2, \dots, u^N\}$

(c) $F : \mathbb{R}^d \to \mathbb{R}^m$

Want to reduce $\mathbb{R}^d$ to $\mathbb{R}^m$, $m \ll d$, such that

$$(1-\delta) \leq \frac{\|F(u^i) - F(u^j)\|_2^2}{\|u^i - u^j\|_2^2} \leq (1+\delta) \tag{1.1}$$

When F is linear, can represent $F(u^i) = Xu$ where $X \in \mathbb{R}^{m \times d}$.
Consider $X \overset{\text{ind.}}{\sim} N(0,1)^{m \times d}$. To scale appropriately, $F(u) = \frac{Xu}{\sqrt{m}}$.
Let $x_i \in \mathbb{R}^d$ be the ith row of X.

$$\implies \left\langle x_i, \frac{u}{\|u\|_2} \right\rangle \sim N(0,1) \tag{1.2}$$

$$\implies Y = \frac{\|Xu\|_2^2}{\|u\|_2^2} = \sum_{i=1}^m \left\langle x_i, \frac{u}{\|u\|_2} \right\rangle^2 \sim \chi_m^2 \tag{1.3}$$

From the concentration bound for χ^2 (see lecture on sub-exponentials), for sufficiently small δ ($\delta \in [0,1)]$)

$$\implies P\left(\left| \frac{1}{m} \frac{\|Xu\|_2^2}{\|u\|_2^2} - 1 \right| \geq \delta \right) \leq 2e^{-\frac{m\delta^2}{8}} \tag{1.4}$$

$$\implies P\left(\frac{\|F(u)\|_2^2}{\|u\|_2^2} \notin [1-\delta, 1+\delta] \right) \leq 2e^{-\frac{m\delta^2}{8}} \tag{1.5}$$

Take a union bound across all $\binom{N}{2}$ pairs of u^i, u^j

$$P\left(\frac{\|F(u^i) - F(u^j)\|_2^2}{\|u^i - u^j\|_2^2} \notin [1-\delta, 1+\delta] \text{ for some } (i,j) \right) \leq \binom{N}{2} 2e^{-\frac{m\delta^2}{8}} \tag{1.6}$$

$$\leq N^2 e^{-\frac{m\delta^2}{8}} \tag{1.7}$$

How to choose m?

$$\varepsilon = P\left(\frac{\|F(u^i) - F(u^j)\|_2^2}{\|u^i - u^j\|_2^2} \notin [1-\delta, 1+\delta] \text{ for some } (i,j)\right) \tag{1.8}$$

$$\varepsilon \leq N^2 e^{-\frac{m\delta^2}{8}} \tag{1.9}$$

$$\implies m \geq \frac{16}{\delta^2} \log \frac{N}{\sqrt{\varepsilon}} \tag{1.10}$$

Memorize: J-L result means that need to choose $m \sim \log N$ to preserve distances.

2 Martingale concentration inequalities

Definition 2.1. Given a martingale Y_k, a *martingale difference sequence* ("innovation" in econ/EE) is

$$D_k = Y_k - Y_{k-1} \tag{2.1}$$

Telescoping the sum, we have

$$Y_n - Y_0 = \sum_{k=1}^{n} D_k \tag{2.2}$$

Notice that

$$\mathbb{E}[D_{k+1} \mid \mathcal{F}_k] = \mathbb{E}[Y_{k+1} \mid \mathcal{F}_k] - \mathbb{E}[Y_k \mid \mathcal{F}_k] = Y_k - Y_k = 0 \tag{2.3}$$

Theorem 2.2 (Wainwright 2.3). *Suppose* $\mathbb{E}[e^{\lambda D_k} \mid \mathcal{F}_{k-1}] \leq e^{\frac{\lambda^2 v_k^2}{2}}$ *for* $|\lambda| < \frac{1}{\alpha_k}$. *Then*

$$P\left(\left|\sum_{k=1}^{n} D_k\right| \geq t\right) \leq \begin{cases} 2e^{-\frac{t^2}{2\sum_{k=1}^{n} v_k^2}}, & \text{if } 0 \leq t \leq \frac{\sum_{k=1}^{n} v_k^2}{\alpha^*} \\ 2e^{-\frac{t}{2\alpha^*}}, & \text{otherwise} \end{cases} \tag{2.4}$$

Exercise 2.3. Prove this

Similar to Hoeffding bound, for martingales we have the following:

Theorem 2.4 (Azuma-Hoeffding). *Suppose* $D_k \in (a_k, b_k)$. *Then*

$$P\left(\left|\sum_{k=1}^{n}\right| \geq t\right) \leq 2e^{-\frac{2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2}} \tag{2.5}$$

This is particularly useful for the *Doob martingale*

$$f(X) = f(X_1, \ldots, X_n), \quad \{X_i\} \text{ ind} \tag{2.6}$$

$$Y_k = \mathbb{E}[f(X) \mid X_1, \ldots, X_k] \tag{2.7}$$

Y_k is a martingale by the tower property.

Example 2.5. Random graph, for $1 \le i < j \le n$ each $X_{ij} \sim \text{Bernoulli}(p)$ represents if edge exists. Let $f(X)$ be the number of cliques in the graph represented by adjacency matrix X. Then Y_k is an *edge-revealing* martingale, and $\mathbb{E}[f(X)]$ is the expected number of cliques in $G(n, p)$.

Lemma 2.6 (Bounded differences)**.** *Let $x, x' \in \mathbb{R}^n$. Define $x_j^{\setminus k} = \begin{cases} x_j, & \text{if } j \neq k \\ x_k', & j = k \end{cases}$.*

Suppose $|f(x) - f(x^{\setminus k})| \le L_k$ (Lipschitz property in Hamming distance). Then

$$P(|f(x) - \mathbb{E}f(x)| \ge t) \le 2e^{-\frac{2t^2}{\sum_{k=1}^n L_k^2}} \tag{2.8}$$

Example 2.7 (U-statistics, $r = 2$)**.**

$$U = \frac{1}{\binom{n}{2}} \sum_{j<k} g(X_i, X_j), \quad \|g\|_\infty \le b \tag{2.9}$$

$$U(x) - U(x^{\setminus k}) \le \frac{1}{\binom{n}{2}} \sum_{j:j\neq k} |g(x_j, x_k) - g(x_j, x_k')| \tag{2.10}$$

$$\le \frac{(n-1)(2b)}{\binom{n}{2}} = \frac{4b}{n} \tag{2.11}$$

$$\implies P(|U - \mathbb{E}U| \ge t) \le 2e^{-\frac{nt^2}{8b^2}} \tag{2.12}$$

Can invert this to get a confidence interval.

Example 2.8 (Clique number in Erdös-Rényi graphs)**.** $G = (V, E)$, $|V| = n$.

$\mathcal{C}(G)$ = cardinality of largest clique ("clique number").

$(i, j) \in E$ with probability p.

Let $X_{ij} = 1_{(i,j)\in E}$, so $\mathcal{C}(G) = f(\{X_{ij}\})$.

$\mathcal{C}(G)$ satisfies a bounded difference; if G' is a graph with one edge added/removed, then $|\mathcal{C}(G) - \mathcal{C}(G')| \le 1$ so

$$P\left(\left|\mathcal{C}(G) - \mathcal{C}(G')\right| \ge \delta\right) \le 2e^{-\delta^2} \tag{2.13}$$

3 Lipschitz functions of Gaussians

Definition 3.1. f is *Lipschitz* if $\exists$ Lipschitz constant $L \ge 0$ such that

$$|f(x) - f(y)| \le L\|x - y\|_2, \quad \forall (x, y) \in \mathbb{R}^n \tag{3.1}$$

Theorem 3.2 (Wainwright 2.4)**.** *Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(0, 1)$. Let $f \sim \text{Lipschitz}(L)$. Then*

$$f(X) - \mathbb{E}f(X) \sim SG(\sigma), \quad \sigma \le L \tag{3.2}$$

$$\implies P(|f(X) - \mathbb{E}f(X)| \ge t) \le 2e^{-\frac{t^2}{2L^2}} \tag{3.3}$$

Hence, any Lipschitz function concentrates in a way that depends on the Lipschitz constant.

TODO: Read this proof before next class, there's a Lemma introducing a trick

TODO: Read Vershynin about geometric functional analysis