Optimización Numérica: Laboratorio 5

rodrigo.mendoza@itam.mx

8 Octubre 2019

1 El Lemma de Johson-Lindenstrauss

Supongamos que tenemos un conjunto de datos $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathbb{R}^n \times \mathbb{R} : i \in [m]\}$ y que n es *muy* grande. Queremos reducir la dimensionalidad de $\mathcal{D}$ preservando la información geométrica en el conjunto de datos (esto es, preservar las distancias relativas $\|x_i - x_j\|$ entre los puntos de datos). En este laboratorio vamos a demostrar una versión del lemma de Johnson-Lindenstrauss.

Lemma 1 (Johnson-Lindenstrauss)**.** *Para todo $\varepsilon \in (0,1)$ y entero m, sea k un entero positivo tal que*

$$k \geq 4\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right)^{-1} \log m \tag{1}$$

Entonces, para todo $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subset \mathbb{R}^n$ existe $f : \mathbb{R}^n \to \mathbb{R}^k$ tal que

$$(1-\varepsilon) \leq \|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1+\varepsilon)\|u - v\|^2 \quad \forall u, v \in S. \tag{2}$$

1. Sean $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ i.i.d., $\mathbf{X} = (X_1, \ldots, X_n)$ y $\mathbf{Y} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$. Sea $\mathbf{\Psi} = [\mathbf{I}_k, \mathbf{0}_{k\times(n-k)}] \in \mathbb{R}^{k\times n}$ con $\mathbf{I}_k \in \mathbb{R}^{k\times k}$ la identidad y $\mathbf{0}_{k\times(n-k)} \in \mathbb{R}^{k\times(n-k)}$ la matriz de ceros. Demuestre que

$$\mathbb{E}[\|\mathbf{\Psi Y}\|^2] = \frac{k}{n} \tag{3}$$

(Pista: Argumente que $X_1^2\|\mathbf{X}\|^{-2}, \ldots, X_n^2\|\mathbf{X}\|^{-2}$ son idénticamente distribuidas y use linealidad de la esperanza.)

2. Vamos a necesitar el siguiente resultado:
Lemma 2. *Sea $k < n$. Entonces,*
   - *Si $\beta < 1$*

$$\Pr\left[\|\mathbf{\Psi Y}\|^2 \leq \beta\frac{k}{n}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \log\beta)\right) \tag{4}$$

   - *Si $\beta > 1$*

$$\Pr\left[\|\mathbf{\Psi Y}\|^2 \geq \beta\frac{k}{n}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \log\beta)\right) \tag{5}$$

   Suponga que $\beta < 1$.

   a) Demuestre que para $t > 0$

$$\Pr\left[\|\mathbf{\Psi Y}\|^2 \leq \beta\frac{k}{n}\right] = \Pr\left[\exp\left\{t\left(k\beta\sum_{i=1}^n X_i^2 - n\sum_{i=1}^k X_i^2\right)\right\} \geq 1\right] \tag{6}$$

   b) Use la *desigualdad de Markov*[1] y el resultado previo para demostrar que,

$$\Pr\left[\|\mathbf{\Psi Y}\|^2 \leq \beta\frac{k}{n}\right] \leq \mathbb{E}\left[\exp\left\{t\left(k\beta\sum_{i=1}^n X_i^2 - n\sum_{i=1}^k X_i^2\right)\right\}\right] \tag{7}$$

[1] https://en.wikipedia.org/wiki/Markov%27s_inequality

c) Si X, Y son variables aleatorias independientes, entonces $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Use este hecho y la función generadora de momentos de la distribución Chi-cuadrada[2] para demostrar que el lado derecho de (7) es igual a

$$g(t) = (1 - 2tk\beta)^{-(n-k)/2}(1 - 2t(k\beta - n))^{-k/2} \tag{8}$$

d) Argumente que para que $g(t)$ tenga sentido neceistamos $t \in (0, (2k\beta)^{-1})$.

e) Demustre que el minimizador de $g(t)$ sujeto a $t \in (0, (2k\beta)^{-1})$ está dado por $t_0 = (1-\beta)(2\beta(n - k\beta))^{-1}$ *(Pista: Demuestre que puede maximizar $g(t)^{-2}$).*

f) Use el resultado anterior para demostrar que

$$\Pr\left[\|\mathbf{\Psi Y}\|^2 \leq \beta\frac{k}{n}\right] \beta^{k/2}\left(1 + \frac{k(1-\beta)}{n-k}\right)^{(n-k)/2} \tag{9}$$

g) Acote el lado derecho de (9) para recuperar (4) *(Pista: Si $t, x > 0$, $e^t \geq (1 + \frac{t}{x})^x$).*

h) Use el mismo procedimiento para demostrar (5).

i) Basado en este Lemma, es probable que $\|\mathbf{\Psi Y}\|^2$ estará lejos de su media?

3. (Ahora demostramos Lemma 1) Suponga que $n > k$ (en otro caso, terminamos). Sea $\mathbf{\Phi} \in \mathbb{R}^{k \times n}$ una transformación lineal que projecta $\mathbb{R}^n$ a un subespacio *aleatorio* de $\mathbb{R}^n$ de dimensión k. Use cualquier de los resultados anteriores para demostrar que:

$$\mathbb{E}\left[\|\mathbf{\Phi}(\mathbf{u} - \mathbf{v})\|^2\right] = \frac{k}{n}\|\mathbf{u} - \mathbf{v}\|^2 \tag{10}$$

Argumente que el Lemma (2) puede ser usado en este caso.

4. Sea $\varepsilon \in (0, 1)$. Use el Lemma (2) para demostrar que para todo $\mathbf{u}, \mathbf{v} \in S$ con $\mathbf{u} \neq \mathbf{v}$,

$$\Pr\left[\|\mathbf{\Phi}(\mathbf{u} - \mathbf{v})\|^2 \leq (1 - \varepsilon)\frac{k}{n}\|\mathbf{u} - \mathbf{v}\|^2\right] \leq \exp\left(-k\frac{\varepsilon^2}{4}\right) \tag{11}$$

(Pista: $\log(1 - x) \leq -x - x^2/2$ para todo $0 \leq x < 1$).

5. Suponga que k satisface (1). Use (11) para demostrar que

$$\Pr\left[\|\mathbf{\Phi}(\mathbf{u} - \mathbf{v})\|^2 \leq (1 - \varepsilon)\frac{k}{n}\|\mathbf{u} - \mathbf{v}\|^2\right] \leq \frac{1}{m^2} \tag{12}$$

6. (Opcional) Siga el mismo razonamiento para demostrar que

$$\Pr\left[\|\mathbf{\Phi}(\mathbf{u} - \mathbf{v})\|^2 \geq (1 + \varepsilon)\frac{k}{n}\|\mathbf{u} - \mathbf{v}\|^2\right] \leq \frac{1}{m^2} \tag{13}$$

(Pista: Use el Lemma 2 y que $\log(1 + x) \leq x - x^2/2 + x^3/3$).

7. Use (12) y (13) para proponer una $f : \mathbb{R}^n \to \mathbb{R}^k$ tal que para $\mathbf{u}, \mathbf{v} \in S$ con $\mathbf{u} \neq \mathbf{v}$

$$\Pr\left[\frac{\|f(\mathbf{u}) - f(\mathbf{v})\|^2}{\|\mathbf{u} - \mathbf{v}\|^2} \notin [(1 - \varepsilon), (1 + \varepsilon)]\right] \leq \frac{2}{m^2} \tag{14}$$

8. Use la *desigualdad de Boole*[3] para demostrar que

$$\Pr\left[\exists \mathbf{u}, \mathbf{v} \in S : \mathbf{u} \neq \mathbf{v} \text{ y } \frac{\|f(\mathbf{u}) - f(\mathbf{v})\|^2}{\|\mathbf{u} - \mathbf{v}\|^2} \notin [(1 - \varepsilon), (1 + \varepsilon)]\right] \leq 1 - \frac{1}{m}$$

Pista: Re-escriba el evento $\{\exists \mathbf{u}, \mathbf{v} \in S : \mathbf{u} \neq \mathbf{v}$ y $\frac{\|f(\mathbf{u}) - f(\mathbf{v})\|^2}{\|\mathbf{u} - \mathbf{v}\|^2} \notin [(1-\varepsilon), (1+\varepsilon)]\}$ como una unión de $\binom{m}{2}$ eventos.

9. ¿ El resultado anterior implica Lemma 1? *(Pista: Sí)*

[2] https://en.wikipedia.org/wiki/Moment-generating_function
[3] https://en.wikipedia.org/wiki/Boole%27s_inequality

2 Aplicación: Programación Lineal

Diseñe e implemente un experimento numérico para demostrar de manera empírica si el tiempo de ejecución de Programación Lineal mejora al reducir la dimensionalidad de los datos con el Lemma de Johnson-Lindenstrauss. Algunas sugerencias:

1. Considere problemas de la forma

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \tag{15}$$

2. Resuelva con el solver `GLPK` de `cvxopt` https://cvxopt.org/userguide/coneprog.html#optional-solvers

3. Reduzca la dimensionalidad de los datos usando el *proyector de Achlioptas* $\mathbf{X} \in \mathbb{R}^{k\times m}$ con entradas tal que $\Pr[\mathbf{X}_{i,j} = +1] = \Pr[\mathbf{X}_{i,j} = -1] = \frac{1}{6}$ y $\Pr[\mathbf{X}_{i,j} = 0] = \frac{2}{3}$.

4. Use $k = \frac{1.8}{\varepsilon^2}\log m$, $\varepsilon = 0.2$.