Luis Serrano
CSE-5095-005

April 22, 2014
Prof. Sanguthevar Rajasekaran

Class Notes

Topics in Big Data Analytics

Data Reduction

Theorem: Let S be any set of n points in $\mathbb{R}^d$. Then, $\exists$ a projection f such that $\forall$ u, v $\in S$, $\|u-v\|^2(1-\varepsilon) \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$. We can find such a project in random polytime.

$$f: \mathbb{R}^d \to \mathbb{R}^k \qquad \text{for } k > 4\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{3}\right)^{-1} \ln n$$

Proof: Consider a random unit vector $Y = \frac{1}{\|X\|}(X_1,\ldots,X_d)$ where $X_i = N(0, 1)$, all independent. Let $Z = \frac{1}{\|X\|}(X_1,\ldots,X_k)$. The expected length of $Z = k/d = \mu$.

Claim: $\mathrm{P}\left[\|Z\| \leq \frac{\beta k}{d}\right] \leq e^{\frac{k}{2}(1-\beta+\ln\beta)}\ \forall\, \beta < 1$ and $\mathrm{P}\left[\|Z\| > \frac{\beta k}{d}\right] \leq e^{\frac{k}{2}(1-\beta+\ln\beta)}\ \forall\, \beta > 1$.

Fact: $\mathrm{E}\left[e^{sX^2}\right] = \frac{1}{\sqrt{1-2s}}$ for any $-\infty < s < ½$, $X = N(0, 1)$.

Fact: (Markov's inequality) If X is a non-negative random variable with $\mathrm{E}[X] = \mu$ then $\mathrm{P}[X \geq a\mu] \leq a^{-1}$ $\Rightarrow \mathrm{P}[X \geq 1] \leq \mu$.

Proof: $\mathrm{P}\left[d\left(X_1^2+\cdots+X_k^2\right) \leq \beta k\left(X_1^2+\cdots+X_d^2\right)\right] = \mathrm{P}\left[\beta k\left(X_1^2+\cdots+X_d^2\right) - d\left(X_1^2+\cdots+X_k^2\right) \geq 0\right] = \mathrm{P}\left[e^{t\left(\beta k\left(X_1^2+\cdots+X_d^2\right)-d\left(X_1^2+\cdots+X_k^2\right)\right)} \geq 1\right]$.

By Markov's inequality the previous probability is $\leq \mathrm{E}\left[e^{t\left(\beta k\left(X_1^2+\cdots+X_d^2\right)-d\left(X_1^2+\cdots+X_k^2\right)\right)}\right]$. We convert it to,

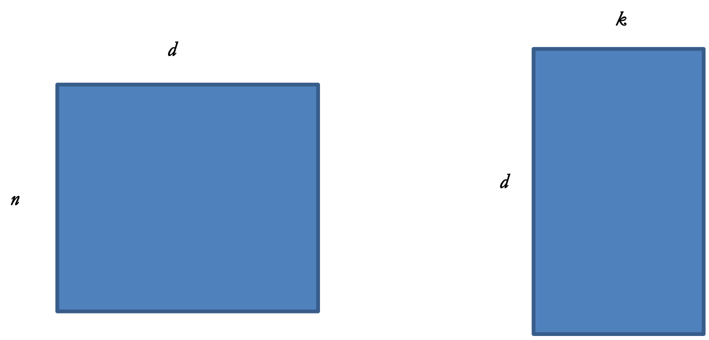
$$\mathrm{E}\left[\left(e^{tk\beta X^2}\right)^{d-k}\left(e^{t(\beta k-d)X^2}\right)^k\right] = \left(\frac{1}{\sqrt{1-2tk\beta}}\right)^{d-k}\left(\frac{1}{\sqrt{1-2t(\beta k-d)}}\right)^k \text{ where } tk\beta < ½.$$

To get the best probability we differentiate it with respect to t, equate it to zero, then we substitute this value of t in it and simplify.

Let $\beta = (1-\varepsilon)$. Then, $\mathrm{P}\left[\|Z\| \leq (1-\varepsilon)\frac{k}{d}\right] \leq \frac{1}{n^2}$ and $\mathrm{P}\left[\|Z\| \geq (1-\varepsilon)\frac{k}{d}\right] \leq \frac{1}{n^2}$. Then, $\mathrm{P}\left[(1-\varepsilon)\frac{k}{d} \leq \|Z\| \leq (1-\varepsilon)\frac{k}{d}\right] \leq \frac{2}{n^2}$. $\Rightarrow$ For a fixed pair, the distance between them is more than a factor β away from their distance in $\mathbb{R}^d$ is $\leq \frac{2}{n^2}$ $\Rightarrow$ The probability that this happens for at least one pair is $\leq \binom{n}{2}\frac{2}{n^2} = 1-\frac{1}{n}$. Repeat this process $\alpha n \ln n$ times. Then, the probability of failure in all of them is $\leq \left(1-\frac{1}{n}\right)^{\alpha n \ln n} \leq \left[\left(1-\frac{1}{n}\right)^n\right]^{\frac{1}{n}\alpha n \ln n} \leq e^{-\alpha \ln n} = n^{-\alpha}$. If $u \in \mathbb{R}^d$ and $u = (X_1, \ldots, X_d)$ then $f(u) = \sqrt{\frac{d}{k}}\left(X_{i_1}, \ldots, X_{i_k}\right)$ where $i_1, \ldots, i_k$ are picked randomly.

<u>Runtime:</u> $O(kn^3 \ln n)$.

Achleoptas Procedure



Data **R**

Consider the above matrices. Each row is a point and each element of R will be $= \begin{cases} \sqrt{3} & \text{with prob } \frac{1}{6} \\ 0 & \text{with prob } \frac{2}{3} \\ -\sqrt{3} & \text{with prob } \frac{1}{6} \end{cases}$

<u>Claim:</u> This projection works with high probability.
<u>Runtime</u>: $O(ndk)$

Learning Algorithm

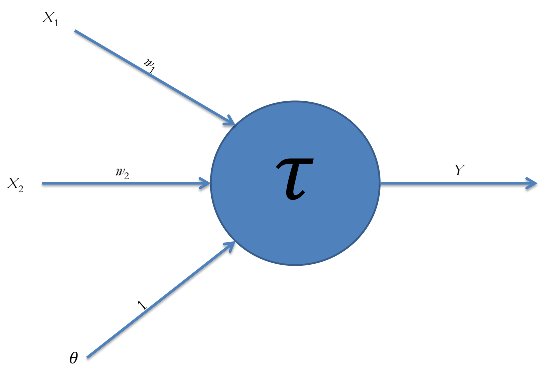
Given positive and negative examples the goal is to learn a concept.

Performance Criteria:

- Sample Complexity □ # of examples
- Time Complexity □ Learning time

Artificial Neural Networks:

It's a directed graph $G(V, E)$ where V are the processors and E are edges with weights.



$Y = 0 \quad$ if $w_1X_1 + w_2X_2 + \Theta < 0$
$Y = 1 \quad$ otherwise