# Homework Assignment 1

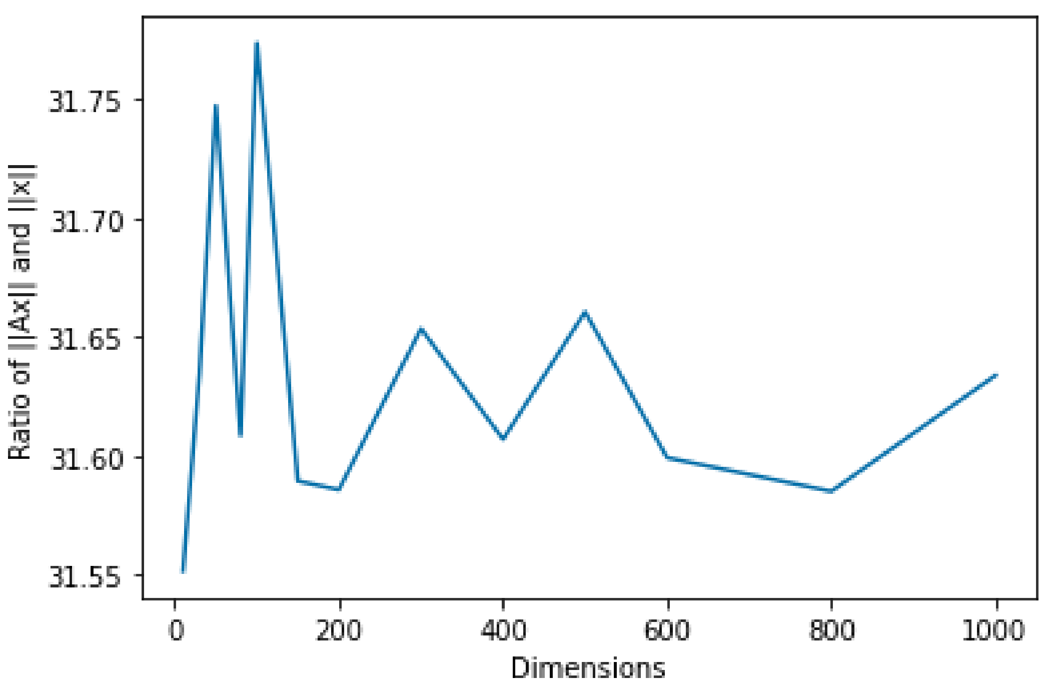
Avirat Belekar
CWID 10454332

October 2, 2020

## 1 Problem 1: Random Projection for Nearest Neighbor Search

Projecting a higher dimensional data onto a lower dimensional subspace using a random matrix whose columns have unit length is known as Random Projection. Random Projection is used to reduce the dimensionality of the data.

Let x be a vector with d dimension. Suppose we have to reduce the dimensionality of x from d to k. Random projection adopts a method of introducing a new matrix A with dimension k X d. According to linear algebra, A.x will have dimensions k X 1 and the new form of x is A.x graphicx




The graph above is plot of the ratio of $||Ax||$ and $||x||$ for various dimensions. It implies that the ratio of $||Ax||$ and $||x||$ does not significantly change even if we increase the dimensions from 10 to 1000. Therefore reducing the dimensionality does not sacrifice the performance of the algorithm.

Problem 2 :

Friday, October 2, 2020 2:09 PM

1. $1\left\{\frac{\Sigma y_i}{N} > 0\right\}$

2. Using Chebyshev's inequality to estimate

$$P\left(\left|\frac{\Sigma y_i}{N} - \mathbb{E}[y_i]\right| \geq \epsilon\right) \leq \frac{Var(y_i)}{n\epsilon^2}$$

$$P\left(\left|\frac{\Sigma y_i}{N} - 0.2\right| \geq \varepsilon\right) < \frac{Var(y_i)}{n\varepsilon^2}$$

$$Var(y_i) = \mathbb{E}[y_i^2] - (\mathbb{E}[y_i])^2$$

$y_i = \pm 1$ $\quad = (1\times 0.6 + 1\times 0.4) - 0.04$

$= 0.96$

$$\text{Std.dev}(y_i) = \sqrt{0.96} = 0.9798$$

$$P\left(\left|\frac{\Sigma y_i}{N} - 0.2\right| \geq \varepsilon\right) \leq \frac{0.96}{n\varepsilon^2}$$

$\because$ Confidence of 0.99 is needed.

$$P\left(\left|\frac{\Sigma y_i}{N} - 0.2\right| \geq \varepsilon\right) \leq 0.01$$

Here choosing a tolerance value of being more than one std deviation away

$\therefore\ \varepsilon = 0.9798.$

$$\Longrightarrow \frac{0.96}{n\times 0.9798\times 0.9798} = 0.01$$

$$\Rightarrow n = \frac{1}{0.01} = 100.$$

3. Chernoff

Here $P(Y_i = 1) \leq 0.6$

$\eta = 0.6$

Now,

$$P(\Sigma Y_i \geq (1+\alpha)\eta n) \leq e^{\frac{-\alpha^2\eta n}{3}}$$

from appendix.

$$P\left(\frac{\Sigma y_i}{n} \geq (1+\alpha)\eta\right) \leq e^{-\frac{\alpha^2\eta n}{3}}$$

Here we want $(1+\alpha)\eta$ to be one standard deviation

$(1+\alpha)\eta = 1.1798 \quad \ldots\ldots \quad \mathbb{E}[Y_i] = 0.2$

But $\eta = 0.6$

$$1+\alpha = \frac{1.1798}{0.6} = 1.9663$$

$\alpha = 0.9663$

In this case, again you want confidence of 0.99

$$e^{\frac{-\alpha^2\eta n}{3}} = 0.01$$

$$\exp\left(-\frac{(0.9663)^2\times 0.6\times n}{3}\right) = 0.01$$

$$\Rightarrow -\frac{(0.9663)^2\times 0.6\times n}{3} = 0.01$$

$$n = \frac{4.6051\times 3}{(0.9663)^2\times 0.6} \approx 23$$

d) We see that for the same tolerance level Chernoff's bound given us a tighter bound & thus number of person needed is substantially less.