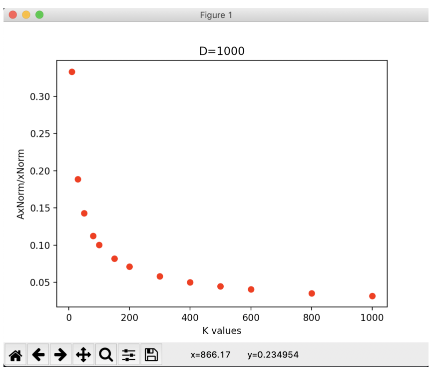
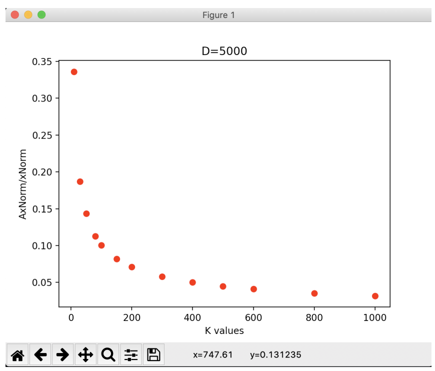
CS 541 Artificial Intelligence: Homework 1

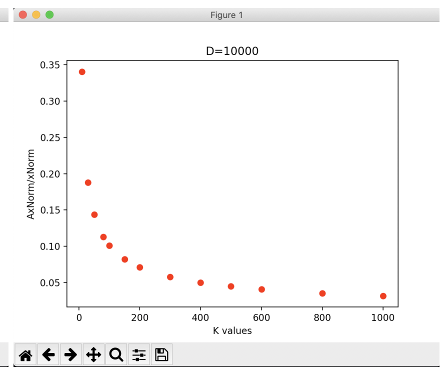
Tarquin Bennett

October 2, 2020

Problem 1: Random Projection for Nearest Neighbor Search







The above images show that when using random projection as the k-values get closer to the size of the original matrix the ratio of $||Ax||$ and $||x||$ approaches zero. This is because as the k-values approach the dimension of the original matrix $||Ax||$ gets closer to $||x||$. When you increase the dimension of the original matrix the resulting ratios and graphs look the same. Therefore, regardless of the original matrix dimension, the smaller you reduce the dimension the higher the ratio of the l_2 distances. This is to be expected since the smaller the dimensions the more data gets compacted and the original l_2 distances get skewed. However, at the lowest k-value the ratio was below 0.5 which means the resulting matrix's data can still be trusted and $||x|| \approx ||Ax||$.

Problem 2: Reliable Data Annotation

1.)

$$Y = \sum_{i=1}^{n} y_i$$

2.)

$$Pr(|Y - 0.2| \geq t) \leq \frac{0.96n}{t^2}$$

$$\frac{0.96n}{t^2} = 0.01, 0.2n - t = 1$$

$$0.01t^2 - 0.96n = 0, t = 0.2n - 1$$

$$0.01(0.2n - 1)^2 - 0.96n = 0$$

$$(0.002n - 0.01)^2 - 0.96n = 0$$

$$4 \times 10^{-6}n^2 - 0.96004n + 0.0001 = 0$$

$$n \approx 2410$$

3.)

$$Pr(Z \geq (1+\alpha)(0.8)n) \leq e^{-\frac{\alpha^2(0.8)n}{3}}$$

$$(1+\alpha)(0.8)n = 1,\ e^{-}\frac{\alpha^2(0.8)n}{3} = 0.01$$

$$1+\alpha = \frac{1}{0.8n}$$

$$\alpha = \frac{1}{0.8n} - 1$$

$$e^{-}\frac{(\frac{1}{0.8n}-1)^2(0.8)n}{3} = 0.01$$

$$-\frac{(\frac{1}{0.8n}-1)^2(0.8)n}{3} = \ln(0.01)$$

$$(\frac{1}{0.8n} - 1)(0.8)n = -3\ln(0.01)$$

$$\frac{0.512n^2 - 1.28n + 0.8}{0.64n} = 13.815511$$

$$0.512n^2 - 1.28n + 0.8 = 8.841927n$$

$$0.512n^2 - 10.121927n + 0.8 = 0$$

$$n \approx 20$$

$$Pr(Z \leq (1-\alpha)(0.2)n) \geq e^{-\frac{\alpha^2(0.2)n}{2}}$$

$$(1-\alpha)(0.2)n = 1,\ e^{-}\frac{\alpha^2(0.2)n}{2} = 0.01$$

$$1-\alpha = \frac{1}{0.2n}$$

$$-\alpha = \frac{1}{0.2n} - 1$$

$$\alpha = 1 - \frac{1}{0.2n}$$

$$e^{-}\frac{(1-\frac{1}{0.2n})^2(0.2)n}{2} = 0.01$$

$$-\frac{(1-\frac{1}{0.2n})^2}{2} = \ln(0.01)$$

$$(1 - \frac{1}{0.2n})^2(0.2)n = -2\ln(0.01)$$

$$\frac{0.008n^2 - 0.08n + 0.2}{0.04n} = 9.21034$$

$$0.008n^2 - 0.08n + 0.2 = 0.368414n$$

$$0.008n^2 - 0.448414n + 0.2 = 0$$

$$n \approx 56$$

4.) The Chebyshev's inequality gives a much broader range of n. That inequality tells us that we need 2410 workers to correctly label an image with a confidence of 0.99. While Chernoff bound, gives us a much tighter region. This bound tells us that we need between 20 and 56 workers to correctly label and image with a confidence of 0.99. This shows that Chernoff bound is a much better estimator then Chebyshev. This allows us to get closer to the correct amount of workers needed to acquire the needed high-quality data.