# Assignment 1.2

Thursday, February 1, 2024 7:10 PM

2. I. $Z = |X - Y|^2$

$= |X^2 - 2XY + Y^2|$

$E[Z] = E[X^2 - 2XY + Y^2|$

$= E[X^2] - E[2XY] + E[Y^2]$

$= E[X^2] - 2E[X]E[Y] + E[Y^2]$

Since X and Y are uniformly distributed, we can represent their expectation as an integral over their domain

$E[X^2] = \int_0^1 x^2 dx = [\frac{1}{3}x^3]_0^1 = \frac{1}{3}$

$E[X] = \int_0^1 x\,dx = \frac{1}{2}$

$E[Z] = \frac{1}{3} - 2(\frac{1}{2})(\frac{1}{2}) + \frac{1}{3}$

$E[Z] = \frac{1}{6}$

Variance of Z: $E[Z^2] - E[Z]^2$

$Z^2 = (X-Y)^4 = X^4 - 4X^3Y + 6X^2Y^2 - 4XY^3 + Y^4$

$E[X^4] = \int_0^1 x^4 dx = [\frac{1}{5}x^5]_0^1 = \frac{1}{5}$

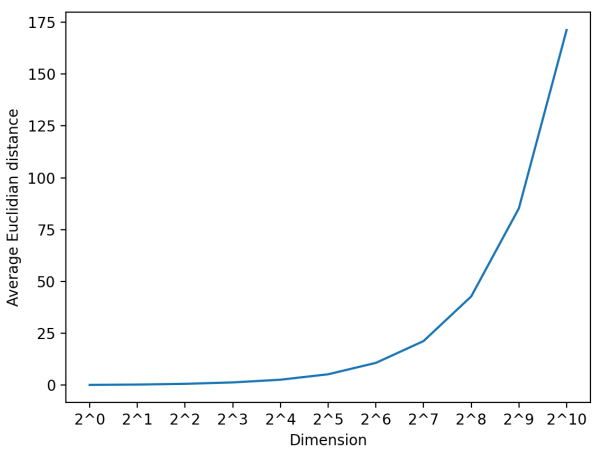
$E[X^3] = \int_0^1 x^3 dx = [\frac{1}{4}x^4]_0^1 = \frac{1}{4}$

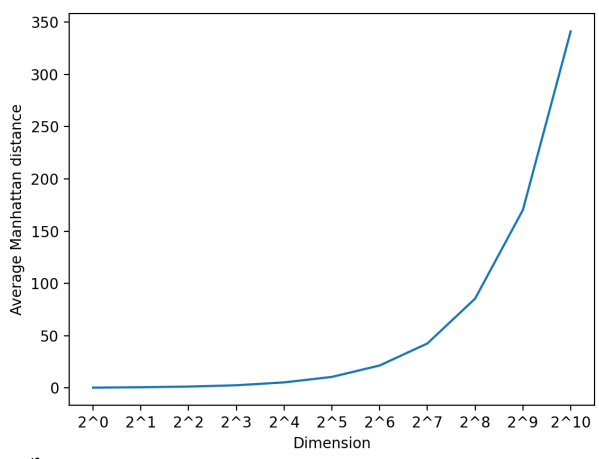
$E[Z^2] - E[Z]^2 = E[X^4 - 4X^3Y + 6X^2Y^2 - 4XY^3 + Y^4] - (\frac{1}{6})^2$

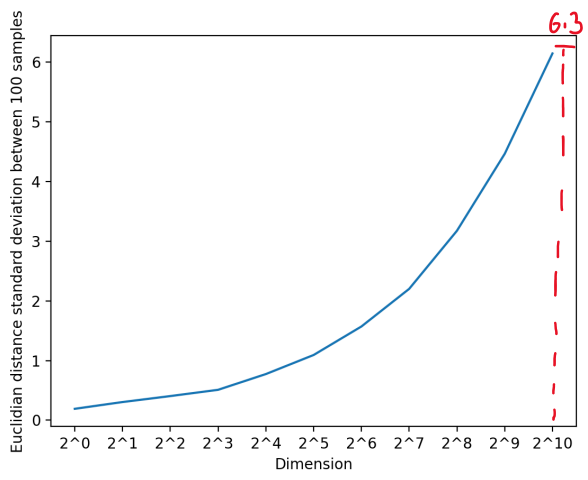
$= \frac{1}{5} - 4(\frac{1}{4})(\frac{1}{2}) + 6(\frac{1}{3})(\frac{1}{3}) - 4(\frac{1}{2})(\frac{1}{4}) + \frac{1}{5} - (\frac{1}{6})^2$

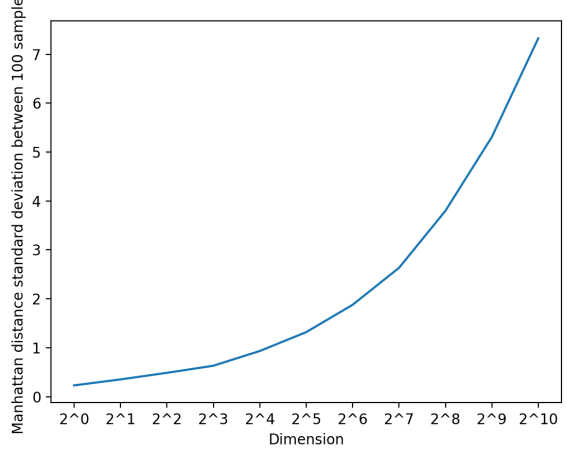
$VAR[Z] = \frac{7}{180}$

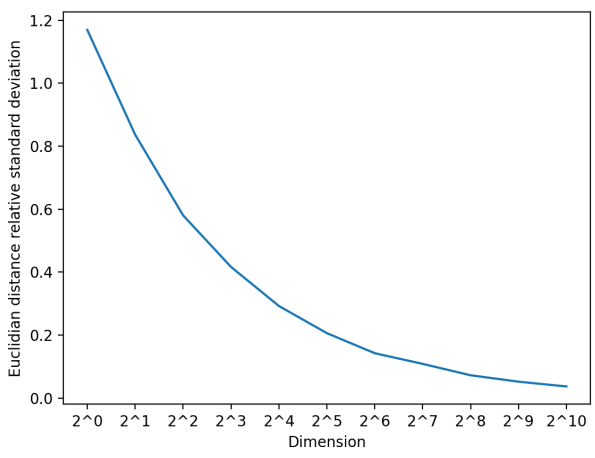
II Code in q_2.py

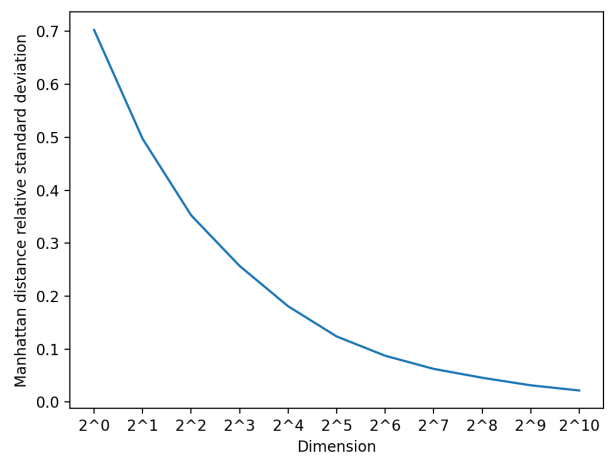



Both the avg distance and the standard deviation increases as the dimension increases. However, the relative standard deviation (SD/AVG) exponentially decreases, meaning that the distances between the points are decreasing as dimension increases.

III. Let $R = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_d - y_d)^2$

$E[R] = E[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_d - y_d)^2]$

$= E[(x_1 - y_1)^2] + \cdots + E[(x_d - y_d)^2]$

from I $= \frac{1}{6} + \frac{1}{6} + \cdots + \frac{1}{6}$

$E[R] = d(\frac{1}{6})$

$Var[R] = Var[(x_1 - y_1)^2 + \cdots + (x_d - y_d)^2]$

$= Var[(x_1 - y_1)^2] + \cdots + Var[(x_d - y_d)^2]$

from I $= \frac{7}{180} + \cdots + \frac{7}{180}$

$= d(\frac{7}{180})$

As the dimensions increase, both expected value and variance increases linearly. However, they increase by different magnitudes ($\frac{1}{6} > \frac{7}{180}$). Furthermore, since $SD = \sqrt{VAR}$, SD increases a lot slower than expected value. This relationship can be visually and manually verified by the plots in II (ie $d = 2^{10} = 1024$, VAR = 39.82, SD = 6.3, matches plot)

IV. $P(|Z - E[Z]| \geq a) \leq \frac{Var[Z]}{a^2}$

VTS: Probability of a random value, R, being far away (→ distance between 2 points) from its expectation approaches 0 as $d \Rightarrow \infty$

Define R to be a random value, E[R] to be it's expected value, then

$|R - E[R]| \geq a$ represents the event that R is "a" units away from it's expected.

$\Rightarrow P(|R - E[R]| \geq a)$ becomes the probability of the event above

sub eqns from III $\Rightarrow$ Markov's

$P(|R - d(\frac{1}{6})| \geq a] \leq \frac{d(\frac{7}{180})}{a^2}$

→ a can be viewed as proportional to variance as $d \propto VAR = \frac{7}{180}d$

let a be proportional to d $\Rightarrow a = cd$ where c is a constant

$P(|R - d(\frac{1}{6})| \geq cd] \leq d(\frac{7}{180})/c^2d^2 = \frac{7}{180c^2}(\frac{1}{d})$

Since $\lim_{d \to \infty} \frac{7}{180c^2}(\frac{1}{d}) = 0$, the probability of the event where R is a certain distance away from its expected approaches 0 as d increases. Since R is a random value representing the distance between any 2 points we can say that in large dimensions, the probability for that distance to be far away from its expected is low and thus all points are roughly same dist apart □

V. $K = 10$, $n = 1000$ $l^d \simeq k/n = \frac{1}{100}$ $l = d\sqrt{1/100}$

| d: | l: |
|---|---|
| 1 | 0.01 |
| 10 | 0.631 |
| 101 | 0.955 |
| 1000 | 0.9954 |
| 10000 | 0.9999 |

The value of l approaches 1 as dimension increases. This means that all the points eventually becomes evenly spread out, so even the 10 closest points to $\hat{x}$ out of 1000 samples are contain in a hyper-cube that spans the entire domain. This becomes an issue in high dimensions as one can no longer identify the closest neighbours as all points are equally close

This issue cannot be solved by increasing the samples since eventually, l will still converge to 1 → increasing samples also computationally expensive

ex. $n = 1000000$

$l^{d=10000} = \frac{10}{1000000} \Rightarrow l = 0.99885$