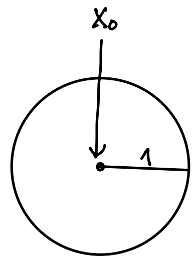
EhS ex. 2.3 Derive Eq. (2.24) pages 22-23

Curse of dimensionality: "all sample points are close to an edge of the sample"



p-dim unitball

N point uniformly distributed in the ball

Want to fit 1NN $\hat{f}(x_0)$ x_0 at origin

What is the median distance from the origin to the closest datapoint?

$\| \; d(p,N)$

then this is the median

$P(\text{closest point has distance} \geq d) = \frac{1}{2}$

Let t_i be the distance from the origin to point i

if the closest point has distance $\geq d$

then all points must have distance $\geq d$

$$P(t_1 \geq d \cap t_2 \geq d \cap \ldots \cap t_N \geq d) = \frac{1}{2}$$

Assume the N points are placed independently

$$P(t_1 \geq d) \cdot P(t_2 \geq d) \cdots P(t_N \geq d) = \frac{1}{2}$$

Now the distribution is uniform in the unit ball so

$$P(t_i \geq d) = 1 - P(t_i < d) = 1 - \frac{\text{volume hypersphere radius } d}{\text{volume hypersphere radius } 1}$$

$\frac{\pi^{p/2}}{(p/2)!} r^p$ according to e.g. wikipedia

(check $p=2$: $\frac{\pi}{1} r^2 = \pi r^2$, ok)

$$P(t_i \geq d) = 1 - \frac{\frac{\pi^{p/2}}{(p/2)!} d^p}{\frac{\pi^{p/2}}{(p/2)!} 1^p} = 1 - d^p$$

$$\prod_{i=1}^{N} P(t_i \geq d) = (1-d^p)^N = \frac{1}{2}$$

$$1 - d^p = \left(\frac{1}{2}\right)^{1/N}$$

$$d^p = 1 - \left(\frac{1}{2}\right)^{\frac{1}{N}}$$

$$d = \left(1 - \left(\frac{1}{2}\right)^{\frac{1}{N}}\right)^{1/p} = d(p, N)$$

the median distance from origin to closest point in unit ball in p dim with N unif. points

→ Extra: make a graph of this!

$N=1,\ p=1 \Rightarrow d = 1 - \frac{1}{2} = \frac{1}{2}$

$N=10,\ p=10 \rightarrow d = \left(1 - \frac{1}{2}^{\frac{1}{10}}\right)^{1/10} = 0.76$

$N=100,\ p=100:\ d = \left(1 - \frac{1}{2}^{\frac{1}{100}}\right)^{\frac{1}{100}} = 0.95$