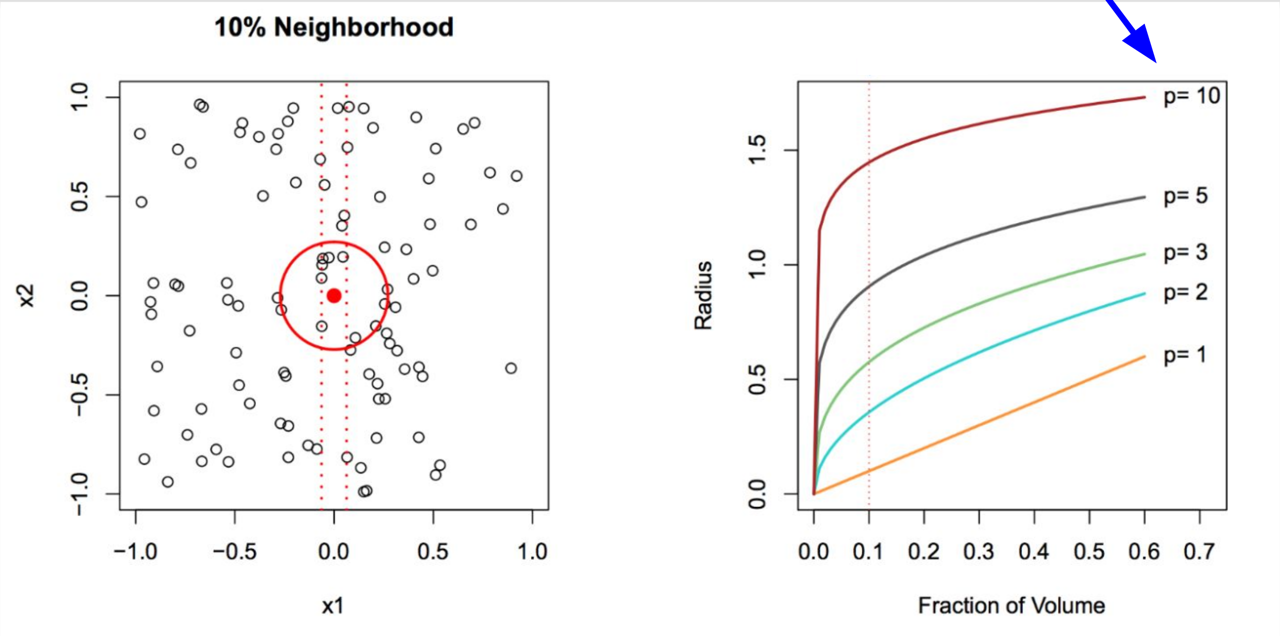
The Curse of Dimensionality

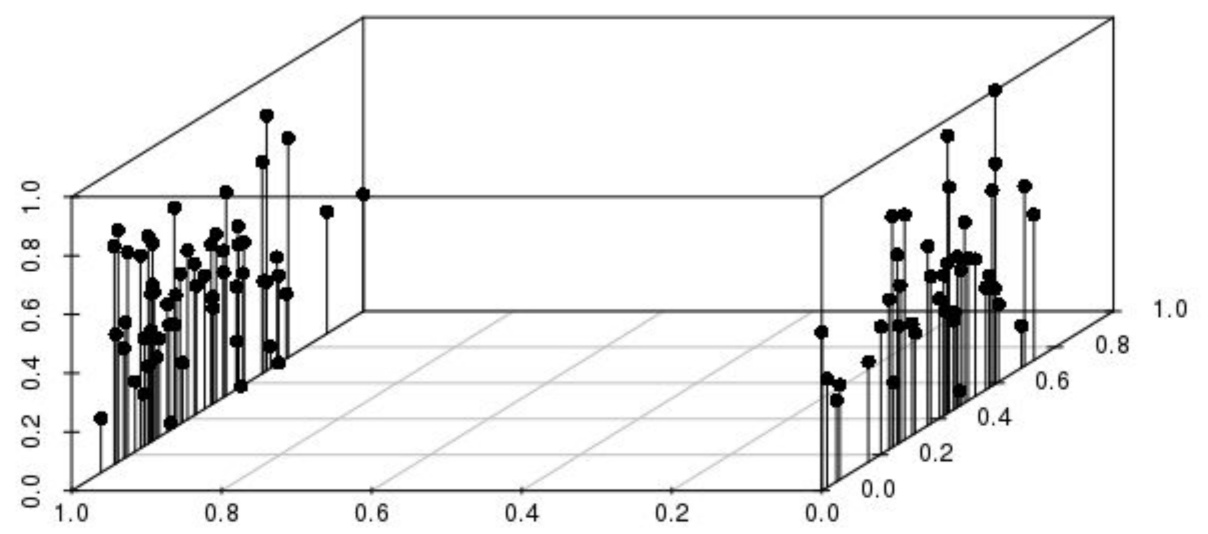
p = dimensionality



When p=1 we are only considering x1. When p=2 we are considering both x1 and x2 and thus adding more ways for datapoints to be farther from each other.

Notice the required radius in 2D is much larger than the required radius in 1D.
As we increase the dimensionality, we lose the concept of locality.

kNN in high dimensions, binary feature (for emphasis)



The Curse of Dimensionality **(another perspective)**

Say you have a dataset with 100 samples, each with only one predictor.

But, one predictor doesn't tell you enough, so you collect a new dataset, and this time you measure 10 predictors for each sample.

How many samples do you need in your new (10 predictor) dataset to achieve the same "sample density" as you originally had (in the one-predictor dataset)?

Just 100^10, that not that many… just

100,000,000,000,000,000,000

"What about for a slightly less extreme changes in d?"

N_0	original number of data points
$V_0 = L^p$	volume of "predictor space" where **L** is an arbitrary (unit) length for the hypercube that our data resides in. **p** is number of predictors
$n_0 = N_0/V_0$	where $\mathbf{n_0}$ is our original data density

Go from **p** → **q** (e.g. N_0=1,000 for 3 predictors → N'? for 7 predictors)

How many data points do we need to maintain n'=n₀? *What is the new N' we need with respect to the old N?*

Strategy: Find density for 1-dimension then up-convert to new dimension.

$n' = N'/V'$ where $V'=L^q$ [← note **p** has been replaced with new number of predictors **q**]

$n_{1\text{-}d} = (N_0/V_0)^{1/p}$

$n' = n_{1\text{-}d}^{\,q} = (N_0/V_0)^{q/p} = N_0^{q/p}/L^q = \boxed{N_0^{q/p}/V'}$

i.e. for new number of parameters q, N needs to be $N_0^{q/p}$, or in our example, $N_0^{7/3}$ or 10,000,000 data points!

The Curse of Dimensionality **(another perspective)**

Say we have a unit (hyper)cube.

We want to create another (hyper)cube inside the outer cube so that we fill X% of the outer cube.

How long must the edges of the inner cube be?

$$e = f^{1/p}$$

p = dimensionality