Problem 1: The curse of dimensionality

(a)

Describe the curse of dimensionality.

For a given sample size, additional features lead to worse rather than better performance.

Why does it make learning difficult in high dimensional spaces?

When we keep adding features, the feature dimensions will keep getting bigger and thinner at the same time. At last, the samples are nearly empty in a high dimensional space, which results in overfitting. The classifier starts learning exceptions that are specific to the training data and do not generalize well when new data is encountered.

As far as I understand, when in high dimensional spaces, there are too many possibilities for classifier to make a suitable classify for samples and the space is too large for samples to teach classifier face every situation. For a new sample, it may be different from any old samples in a high dimensional space, so the classifier do not know how to classify it.

(b)

What is the ratio between the volume of the crust and the volume of the hypersphere? Using Math Type edit

$$\text{ratio} = \frac{V_d(r) - V_d(r-\varepsilon)}{V_d(r)}$$

$$= \frac{\frac{r^d \pi^{d/2}}{\Gamma(d/2+1)} - \frac{(r-\varepsilon)^d \pi^{d/2}}{\Gamma(d/2+1)}}{\frac{r^d \pi^{d/2}}{\Gamma(d/2+1)}}$$

$$= 1 - \frac{(r-\varepsilon)^d \pi^{d/2}}{r^d \pi^{d/2}}$$

$$= 1 - (\frac{r-\varepsilon}{r})^d$$

How does the ratio change as d increases?

because $r-\varepsilon < r$, $\frac{r-\varepsilon}{r}$ <1, when d increases, $(\frac{r-\varepsilon}{r})^d$ decrease, ratio increases

(c)

(6 points) We assume that N data points are uniformly distributed in a 100-dimensional unit hypersphere (i.e. $r = 1$) centered at the origin, and the target point x is also located at the origin. Define a hyperspherical neighborhood around the target point with radius r'. How big should r' be to ensure that the hypersperical neighborhood contains 1% of the data (on average)? How big to contain 10%?

Because N data points are uniformly distributed, hyperspherical neighborhood contains 1% equals to the volume of r' is 1% volume of r.

$$\frac{\frac{(r')^d \pi^{d/2}}{\Gamma(d/2+1)}}{\frac{r^d \pi^{d/2}}{\Gamma(d/2+1)}} = 1\%$$

$$(\frac{r'}{r})^d = \frac{1}{100}$$

$$r = 1, d = 100$$

$$r' = \sqrt[-100]{100} \approx 0.955$$

When contain 10%,

$$(\frac{r'}{r})^d = \frac{1}{10}$$

$$r' = \sqrt[-100]{10} \approx 0.977$$