Problem 4

Tuesday, June 11, 2024 4:29 PM

Curse of dimensionality

Parametric approaches often perform badly when p is large

a) $\frac{1}{10}$ since 10%

b) $\frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100}$ = 1% of the data

c) $\left(\frac{1}{10}\right)^{100} = \frac{1}{10^{100}} = \left(.10^{100}\right)\%$

d) since our window of data is so small that is very few points in the neighborhood for training. thus the points selected can be far from the test points

e) length is

L^p = area given p dim.

cube is $10\% = 1/10$

V_c = total volume * fraction observed

$X^1 = 1^1 \cdot \frac{1}{10} = .10$

$X^2 = 1^2 \cdot \frac{1}{100} = \frac{\sqrt{1}}{\sqrt{10^2}} = \frac{1}{10} = .10$

$X^{100} = 1^{100} \cdot \frac{1}{10^{100}} = \frac{\sqrt[100]{1}}{\sqrt[100]{10^{1000}}} = \frac{1}{10} = .10$

However the 10% is always fixed now

$S^p = \frac{1}{10}$ thus $S = \left(\frac{1}{10}\right)^{1/p}$

$p=1 = (.10)^1 = .10$

$p=2 = (.10)^{1/2} = .316$

$p=100 = (.100)^{1/100} = .977$