Lecture: High Dimensionality

next tuesday: recap. Wednesday: lab.

Recap

- MSE
- MAE
- R^2 . MSE & normalize by value of y.

  if it is negative: you're peforming worse than just taking the MSE.

Calibration: shows that we are close to the true value.

$$\sqrt{E[|E[Y|f(x)] - f(x)||^2]}$$

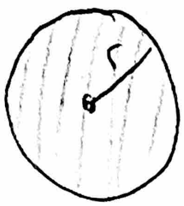
Probabilities: $\sqrt{E[|P[Y|f(x)] - f(x)||^2]}$

example $f(x) = 0.3 \Rightarrow P[Y | f(x) = 0.3]$

30% of those my classifier says should be spam, when I look at my subset there should be 30% spam.

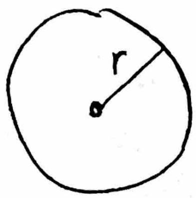
Y is the true label. Y depends on X & as X is dependent on f(x) y is dependent on f(x) as well.

High dimension



Ball of radius r

B_r ball $r=1$ unit ball



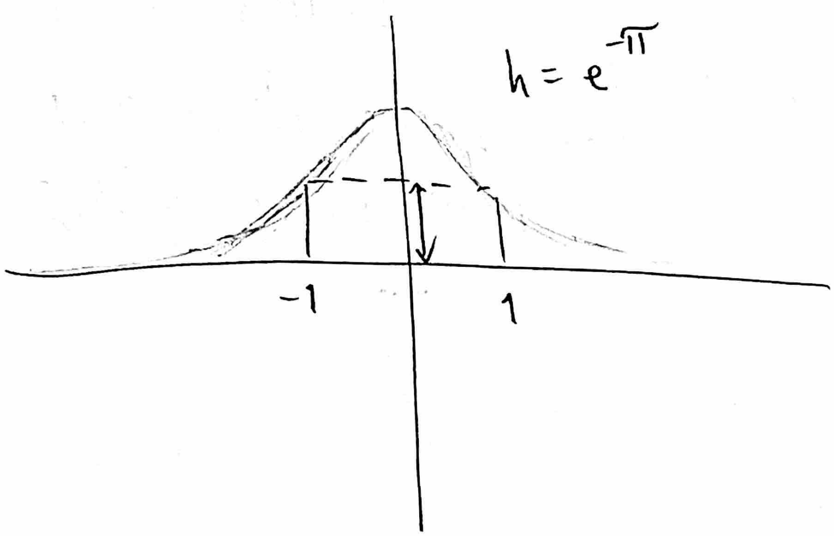
S_r sphere $r=1$ unit sphere

S_r is just the shell. The surface of the ball. The inside of the ball.

y : a RV in $\mathbb{R}^d$

density $f(x)$. $f(x) = e^{-\pi|x|^2}$ $x \in \mathbb{R}^d$

is a normalized Gaussian



if I take y as before what is the probability that y is within the unit ball?

e.g. $\mathbb{P}(y \in B_1) = \int_{B_1} e^{-\pi \cdot |x|^2} \, dx$ ← requirement that $x < 1$

$\geq e^{-\pi} \underbrace{|B_1|}$

this probability is an upper bound to the volume.

Code: take regular gaussian & multiply by

Result:
Start with prob close to 1, but the probability decreases to 0 as we increase the dimensions.

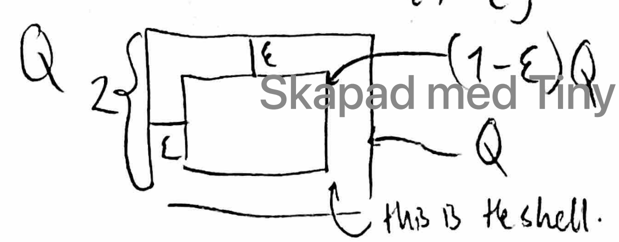
↑ The blue line.

[As we increase the dimensions, the volume of the ball goes to 0. We never land in the ball as we scale up dimensions.]

↑ with our probabilities Pts.

Take $Q = [-1, 1]^d$ and $\varepsilon \in (0, 1)$

$$\underbrace{(1-\varepsilon)Q}_{\text{scaling } Q \text{ down with } (1-\varepsilon)} = [-(1-\varepsilon), (1-\varepsilon)]^d$$



$|Q| = 2^d$

$|(1-\varepsilon)Q| = (1-\varepsilon)^d 2^d$

$$\frac{|\overbrace{Q \setminus (1-\varepsilon)Q}^{\text{the shell}}|}{|Q|} = 1-(1-\varepsilon)^d$$

High dimension: don't trust your instincts.

30... 20...

1 1 1 1 1 0 0

Take $\varepsilon = 1/d$ $\quad 1-(1-1/d)^d \to 1-\frac{1}{e}$

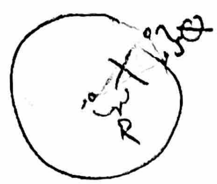
it doesn't matter what ε is, when we increase the dimension, the volume can be approximated by the shell. Most of the volume is close to the edge of the cube.

The volume to the Q goes to infinity in high dim.
The volume of the ball goes to 0 in high dim.

Scaling of dimension

We want a uniform X on B_1

$X = R\Theta$



By symmetry, $\Theta \sim \text{unif.}(S_1)$ of S_1

$F_R(r) = \mathbb{P}(R \le r)$

$= \mathbb{P}(X \in B_r) = \dfrac{|B_r|}{|B_1|}$

$= \dfrac{r^d |B_1|}{|B_1|} = r^d$

Lecture 13.

- inversion sampling technique
- accept-reject sampling

Inversion sampling

$$U \sim \mathrm{Unif}([0,1])$$

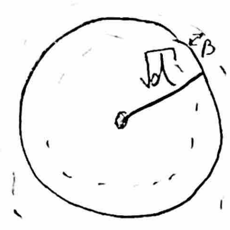
$$F_R^{-1}(U) = U^{1/d} =: R$$

if we have something that is or looks symmetric use the unit ball over Q for our accept-reject sampling (for higher dim). For lower dim. it doesn't matter. But in higher dim it does as Q would reject alot.

The annulus theorem

$$X \sim f$$

$$f = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}|x|^2}$$



$$x = (x_1, \dots, x_d)$$

$$E[X_1^2] = 1$$

$$E[|X|^2] = \sum_{i=1}^{d} E[|X_i|^2] = d$$

Unlikely B O & close to 1 Rather it in a region around

This can be used to do dimensionality reduction.