Density Estimation

E.g.: Assuming tweets from an account are independently generated randomly. Create a robot account that generates more such tweets.

$f(\text{Tweet}) = \text{Score of the tweet}$

wisdomofchopra.com

It has been said by some that the thoughts and tweets of Mr. Chopra are indistinguishable from a set of profound sounding words put together in a random order, particularly the tweets tagged with "#cosmisconciousness". This site aims to test that claim! Each "quote" is generated from a list of words that can be found in Deepak Chopra's Twitter stream randomly stuck together in a sentence.

"A formless void transforms total mysteries"

RECEIVE MORE WISDOM...

Tweet the wisdom

Disclaimer: This is intended for entertainment purposes only. It in no way reflects the thoughts of any real person.

Density Estimation

"A formless void transforms total mysteries"

RECEIVE MORE WISDOM...

Tweet the wisdom

To generate such sentences randomly, we need to be able to assign a probability score to every possible 128 character sentence, giving high scores to those that are likely to be from the original source.

A density estimation model takes in several samples from a random source, and outputs a model that assigns a probability score to every possible instance.

Density Estimation

Σ

sum over
all tweets 10^{8}
{a .. z}

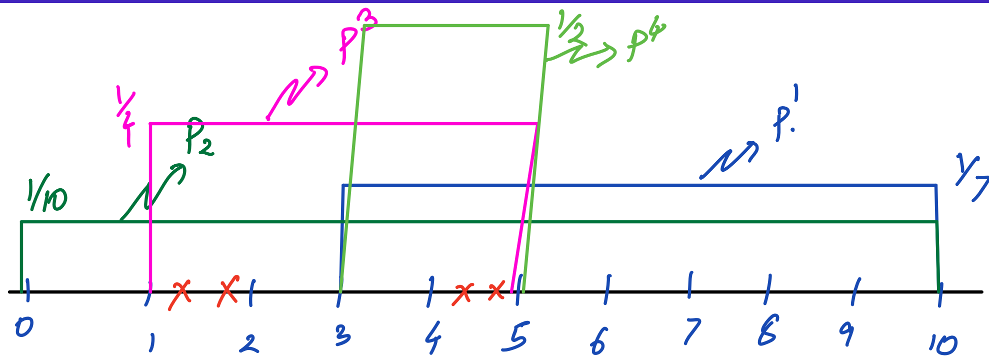
$P(\text{tweet}) = 1$

- Data: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$
- $\mathbf{x}^i \in \mathbb{R}^d$
- Probability mapping $P : \mathbb{R}^d \rightarrow \mathbb{R}_+$ that 'sums' to one.
- Goal : $P(\mathbf{x})$ is large if $\mathbf{x} \in$ Data, and low otherwise.
- Loss $= \frac{1}{n} \sum_{i=1}^{n} -\log\left(P(\mathbf{x}^i)\right)$

$P(x^i)$ is large.

$P(\text{anything}) = 10^{10}$

Density Estimation Illustration 1



$x^1 : [1.2]$

$x^2 : [1.9]$

$x^3 : [4.3]$

$x^4 : [4.8]$

p^1 = Uniform in $[3, 10]$ $\quad 0, 0, \frac{1}{7}, \frac{1}{7}$

p^2 = Uniform in $[0, 10]$ $\quad \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}$

p^3 = Uniform in $[1, 5]$ $\quad \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$

p^4 = Uniform in $[3, 5]$ $\quad 0, 0, \frac{1}{2}, \frac{1}{2}$

$$\text{loss}[p^4] = \text{loss}[p^1] = \infty > \text{loss}[p^2] > \text{loss}[p^3]$$

Density Estimation Illustration 2

$d=2$

$(1.1,\ 1.3)$

$(0.9,\ 0.7)$

$(2.1,\ -1)$

$(5.1,\ 0.1)$

$(2.2,\ -0.9)$

$(5.1,\ 0.0)$

$(0.9,\ 1.2)$

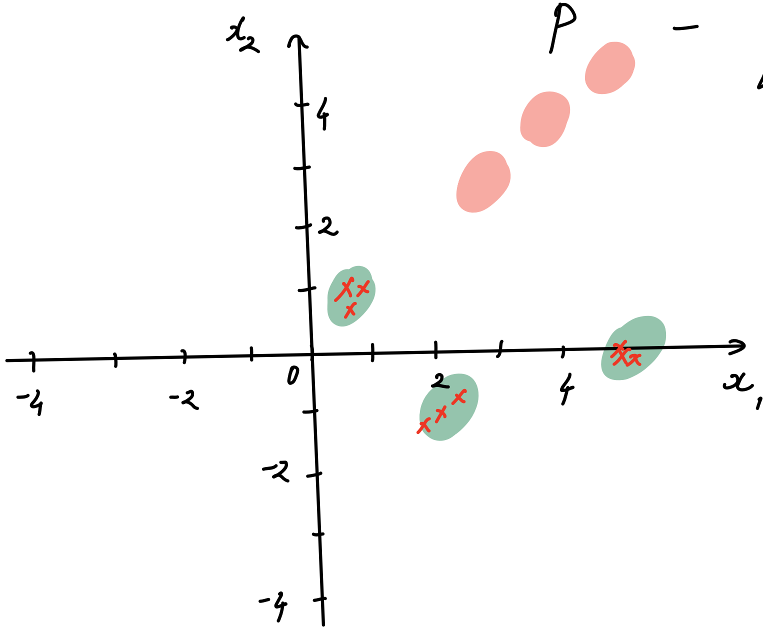
$(1.9,\ -1.1)$

$(4.8,\ -0.1)$

Gaussian Mixture Model

$p^1 - \begin{bmatrix}3\\3\end{bmatrix},\ \begin{bmatrix}5\\5\end{bmatrix}\ \begin{bmatrix}4\\4\end{bmatrix}$

$p^2 - \begin{bmatrix}1\\1\end{bmatrix}\ \begin{bmatrix}2\\-1\end{bmatrix}\ \begin{bmatrix}5\\0\end{bmatrix}$



Clustering