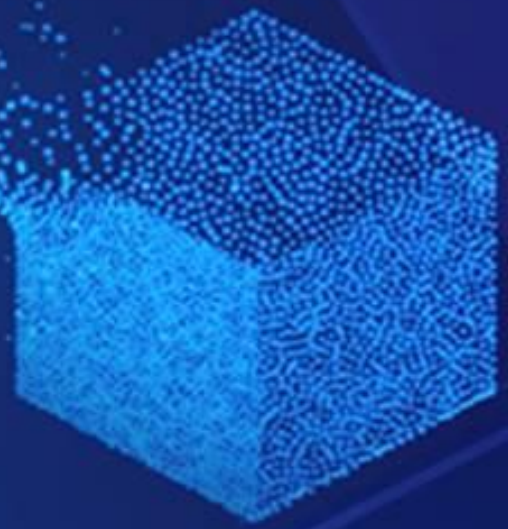
TECH
talk

AGILISIUM
BIG ON CLOUD. BIG ON DATA.



Magic of Models

Venkatesh Vinayakarao

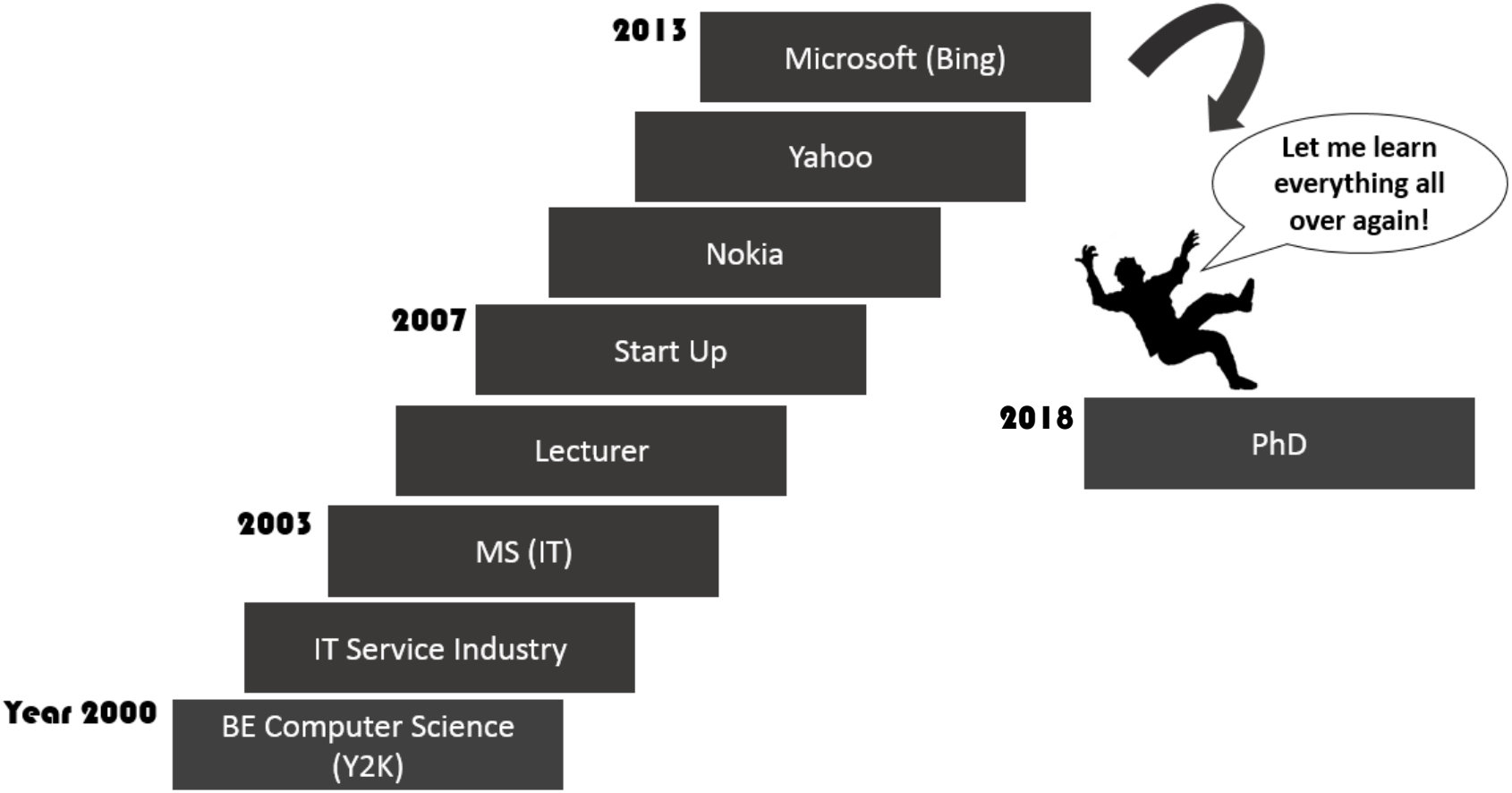
Sep
24
2019

3:00 PM - 4:00 PM

Agilisium Consulting
Oval Conference Room

Agilisium, Rattha Tek Meadows, No.51, Tower B, 6th Floor, OMR, Sholinganallur Chennai- 600119

About Me


2013
Microsoft (Bing)
Yahoo
Nokia
2007
Start Up
Lecturer
2003
MS (IT)
IT Service Industry
Year 2000
BE Computer Science (Y2K)
Let me learn everything all over again!
2018
PhD

Big Data is Ubiquitous

Big Data Landscape

Source: © Matt Turck (@mattturck) and ShivonZilis (@shivonz)

To make sense out of the data…

How to tame complexity?

Answer: Design effective models!

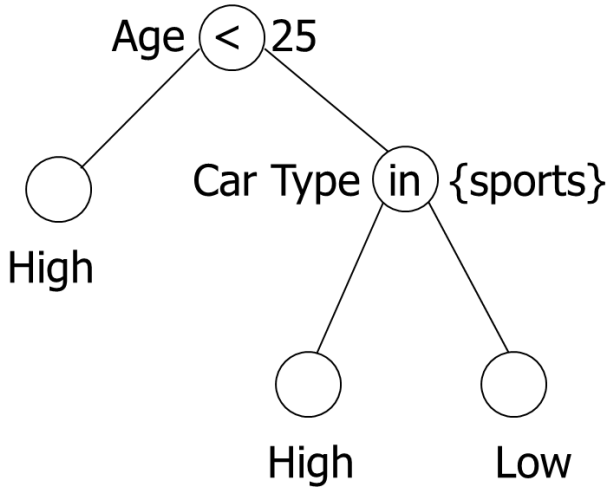
A Simple Model

Data Set

Age	Car Type	Risk
23	Family	High
17	Sports	High
43	Sports	High
68	Family	Low
32	Truck	Low
20	Family	High

Question: What is the risk (high or low) if age is below 25?

Decision Tree



Agenda

- Case Study 1: How to make decisions based on data?
  - Bayesian Data Analysis
- Case Study 2: How to effectively retrieve relevant documents?
  - Vector Space Models

Thomas Bayes, 1701 to 1761

Bayesian Data Analysis and Beta Distribution

Venkatesh Vinayakarao

The Case of Coin Flips

General Assumption: If a coin is fair! Heads (H) and Tails (T) are equally likely. But, coin need not be fair ☹

Experiment with **Coin - 1**
HHHTTTHTHT

Experiment with **Coin - 2**
HHHHHHHHHT

Coin-1 more likely to be fair when compared to **coin-2**.

Our Beliefs

- Can we find a structured way to determine coin's nature?

Coin-1

Data Observed	None	H	T	H	T
Belief	Fair	Skewed	Fair	Skewed	Fair

Coin-2

Data Observed	None	H	H	H	H
Belief	Fair	Skewed	More Skewed	Even More…	Even Even More…

Prior Belief

Belief Updates

Priors

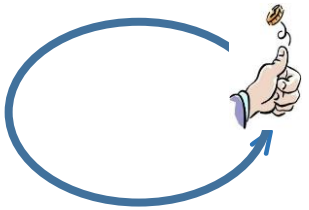
- Priors can be strong or weak

Weak Prior

New Coin → A few observations sufficient to change our belief significantly.

Strong Prior

Coin is lab tested for 1 Million Tosses.
50% H, 50% T observed.



→ One more observation will not change our belief significantly.

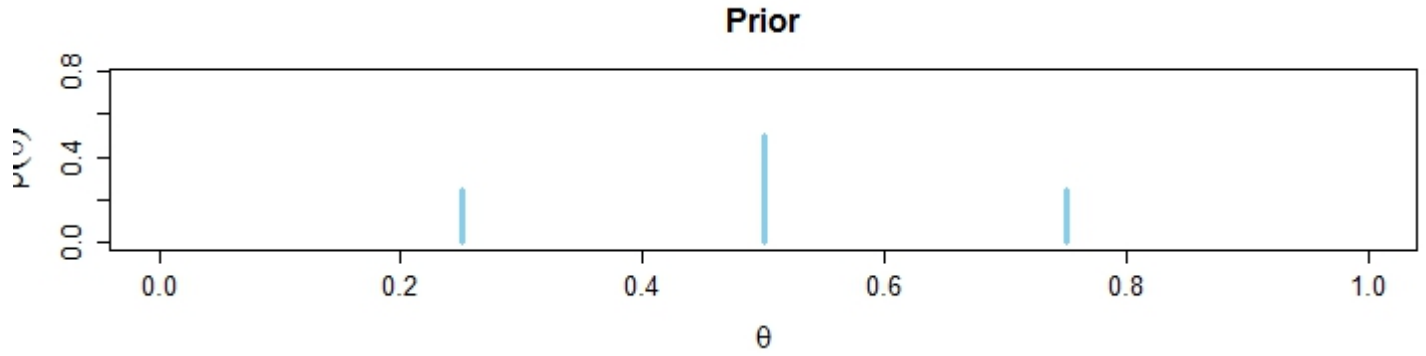
HyperParameter

- Prior probability (of Heads) could be anything:
  - 0.5 → Fair Coin
  - 0.25 → Skewed towards Tails
  - 0.75 → Skewed towards Heads
  - 1 → Head is guaranteed!
  - 0 → Both sides are Tails.

We use θ as a HyperParameter to visualize what happens for different values.

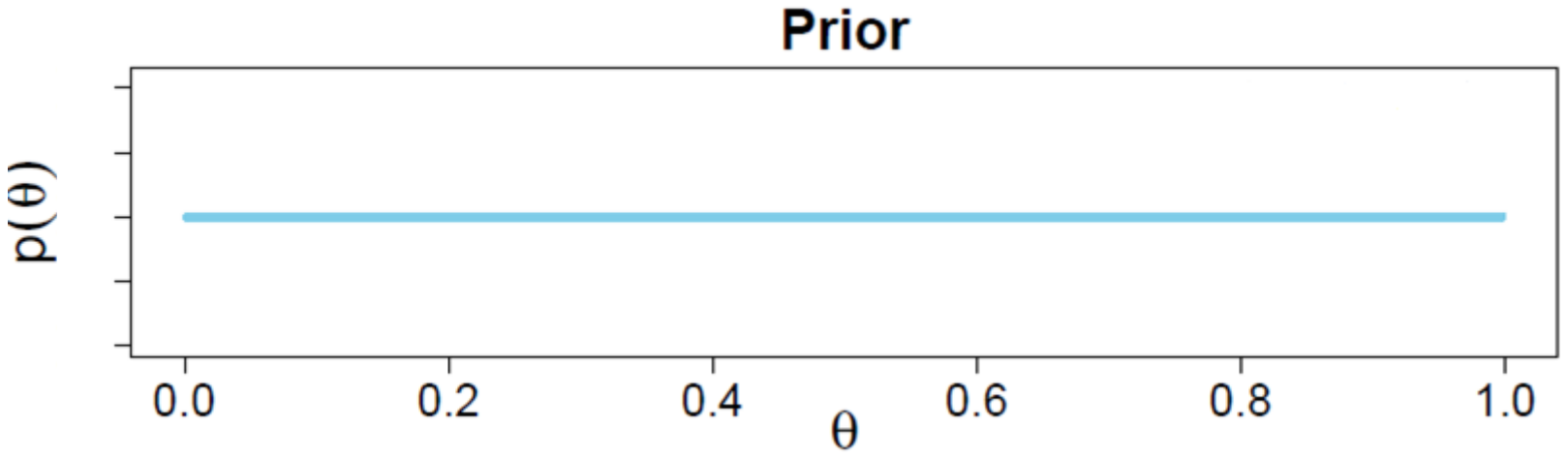
World of Distributions

Discrete Distribution of Prior. Since I typically perceive coins as fair, Prior belief peaks at 0.5.



Another Possibility

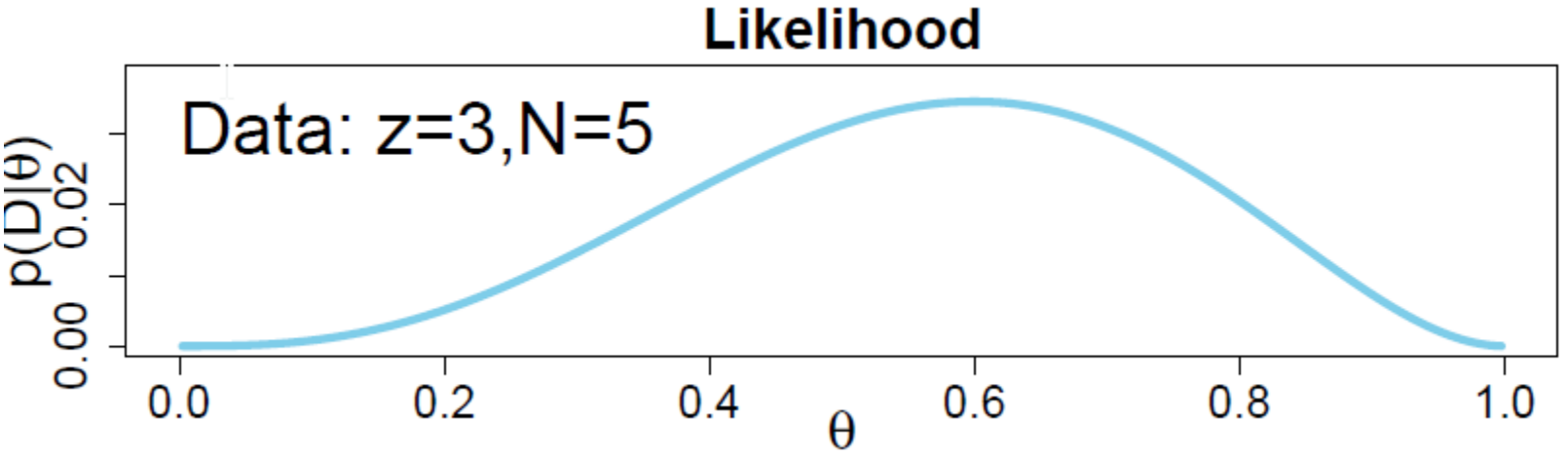
I may also choose to be unbiased! i.e., θ may take any value equally likely.



A Continuous Uniform Distribution!

Observations

Let's flip the coin (N) 5 times. We observe (z) 3 Heads.



Impact of Data

Belief is influenced by Observations. But, note that:

Belief ≠ observation

Bayes' Rule

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\Pr(\theta)}{\Pr(D)}$$

Numerator is easy

- p(θ) was uniform. So, nothing to calculate.
- How to calculate p(D|θ)?

Jacob Bernoulli
1655 – 1705.

$$\theta^z(1-\theta)^{n-z}$$

If D observed is HHHTT and θ is 0.5, We have:

$$p(D|\theta) = (0.5)^3(1-0.5)^{5-3}$$

Remember, two things: 1)we are interested in the distribution 2) Order of H,T does not matter.

Painful Denominator

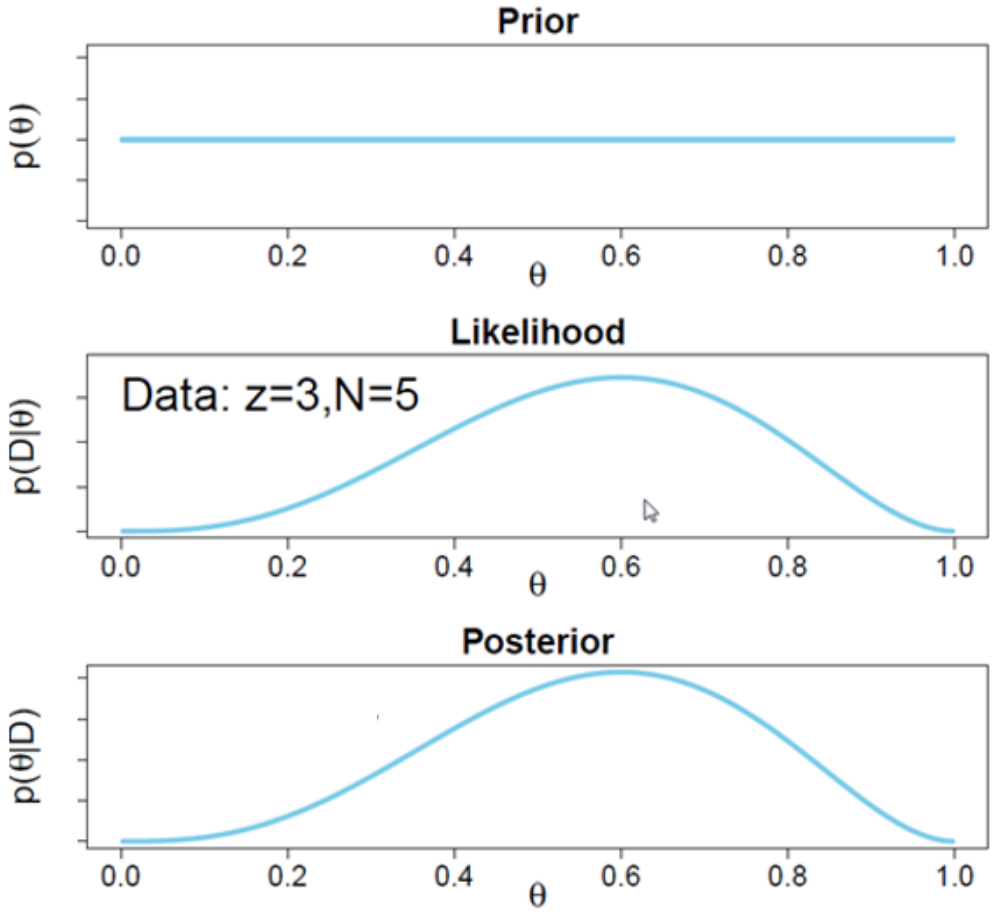
- Recall, for discrete distributions:

$$P(\theta|D) = \frac{P(D|\theta)\,P(\theta)}{\sum_i P(D|\theta_i)\,P(\theta_i)}$$

- And, for continuous distributions:

$$P(\theta|D) = \frac{P(D|\theta)\,P(\theta)}{\int P(D|\theta)\,P(\theta)\,\mathrm{d}\theta}$$

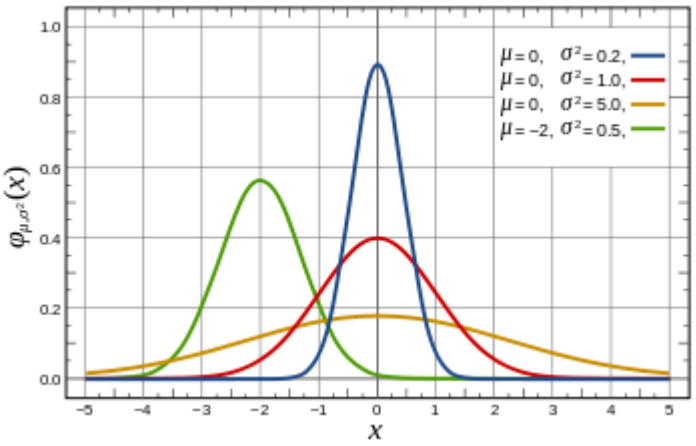
Bayesian Update



A Simpler Way

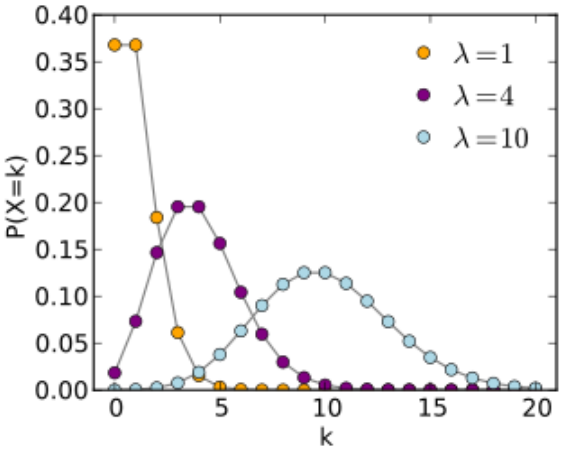
Form, Functions and Distributions

Normal (or Gaussian)



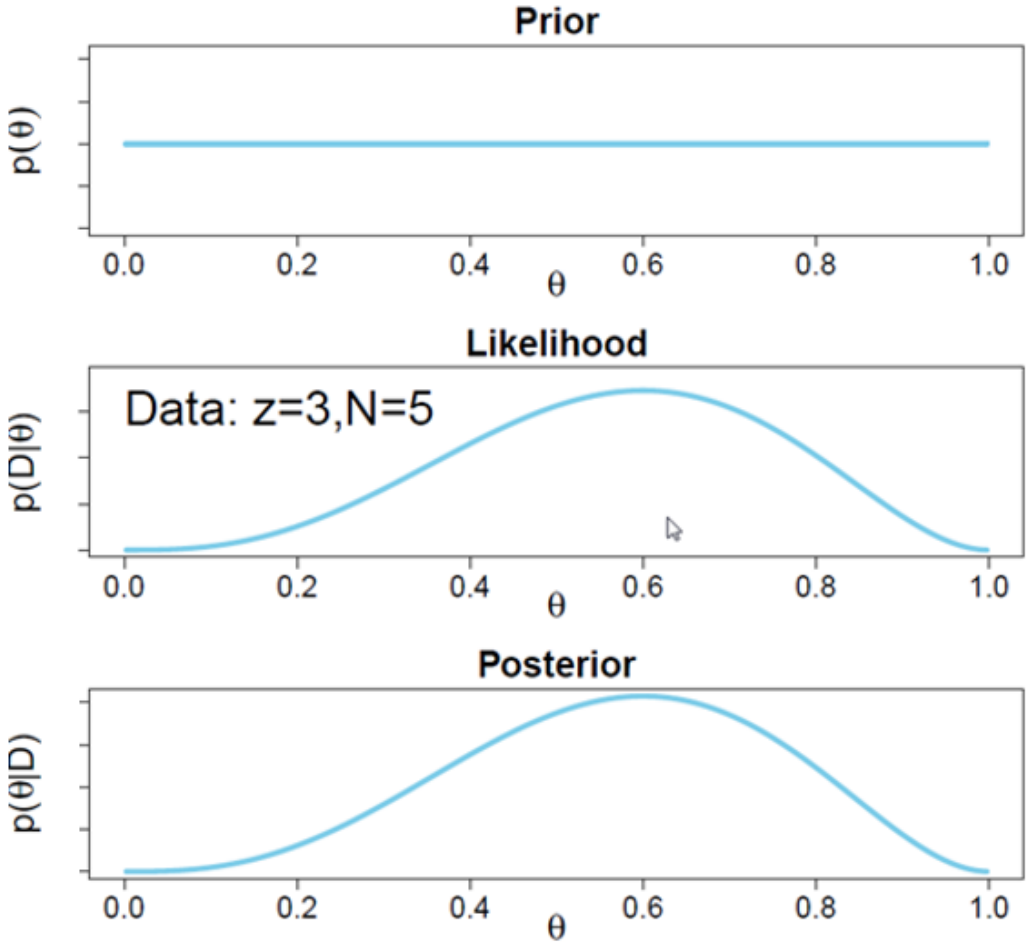
$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Poisson

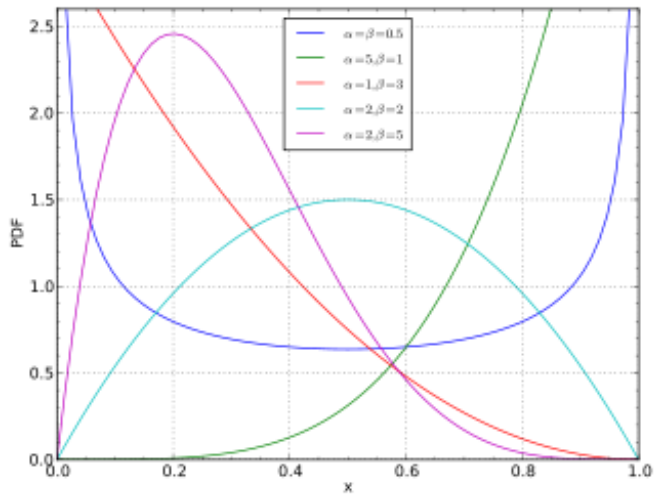


$$f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

What form will suit us?



Beta Distribution

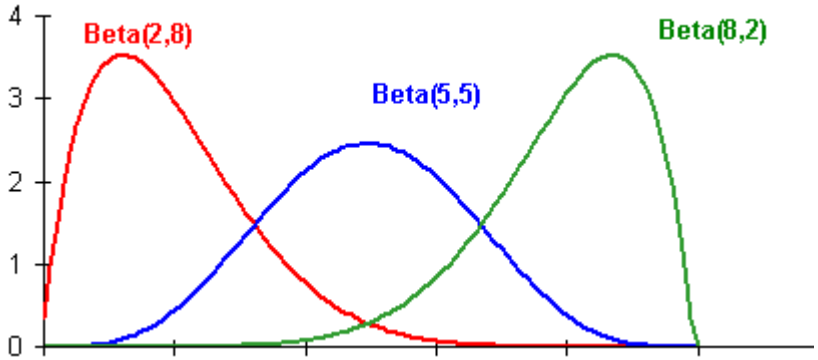


$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

Vadivelu, a famous Tamil comedian. This is one of his great expression - terrified and confused.

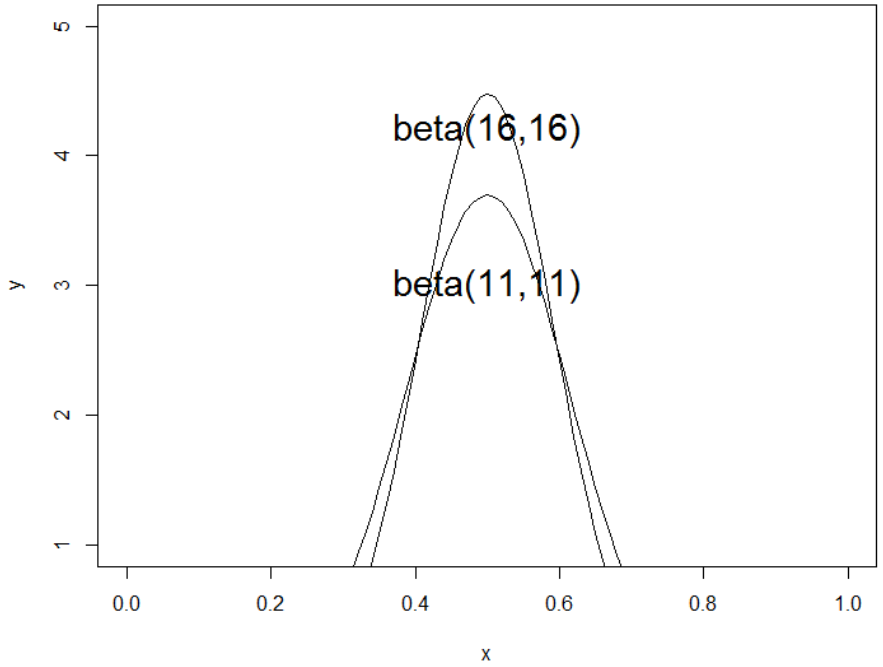
Beta Distribution

- Takes two parameters Beta(a,b)
- For now, assume $a = \#H + 1$ and $b = \#T + 1$.
- After 10 flips, we may end up in one of these three:



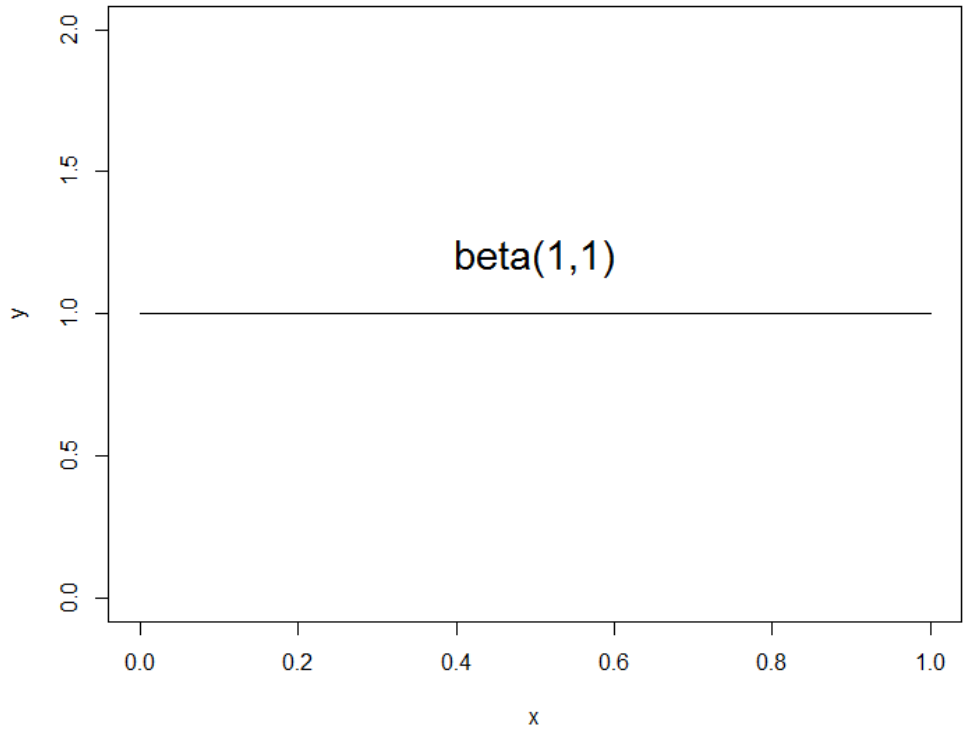
Prior and Posterior

Let's say we have a Strong Prior – Beta (11,11). What should happen if we see 10 more observations with 5H and 5T?



Prior and Posterior…

What if we have not seen any data?



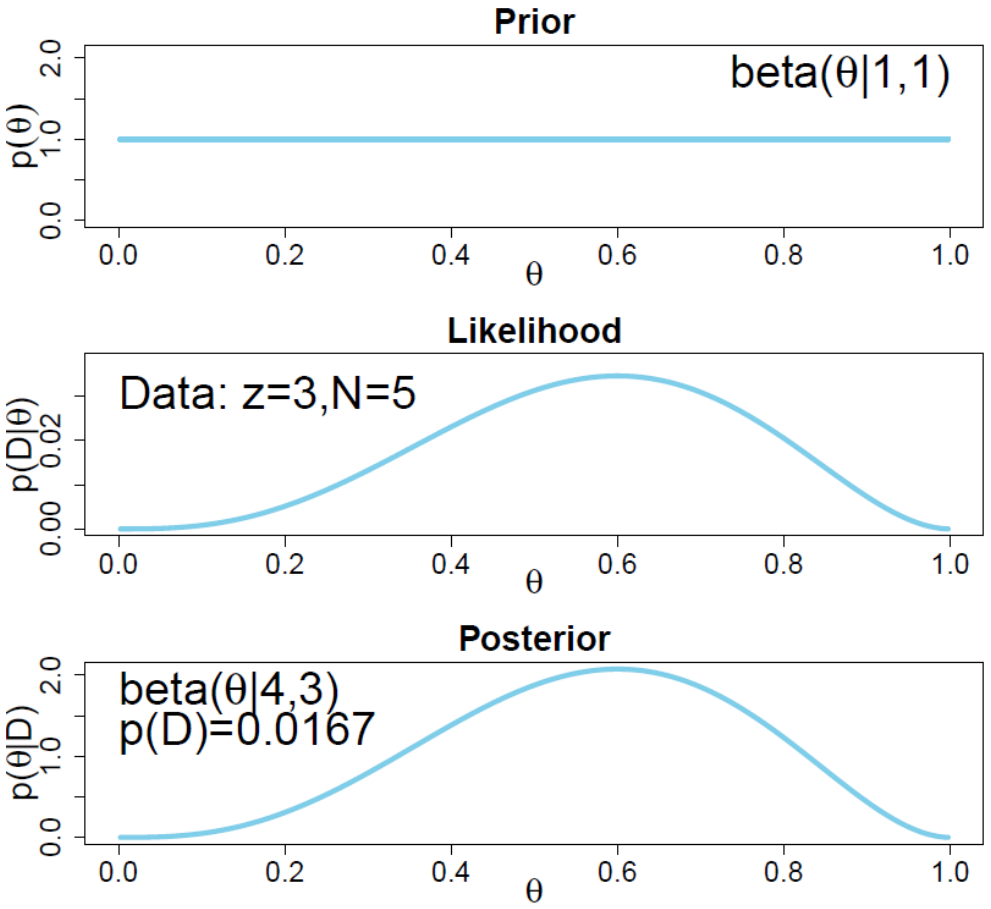
Conjugate Prior

So, we see that:

Such Priors that have same form are called Conjugate Priors

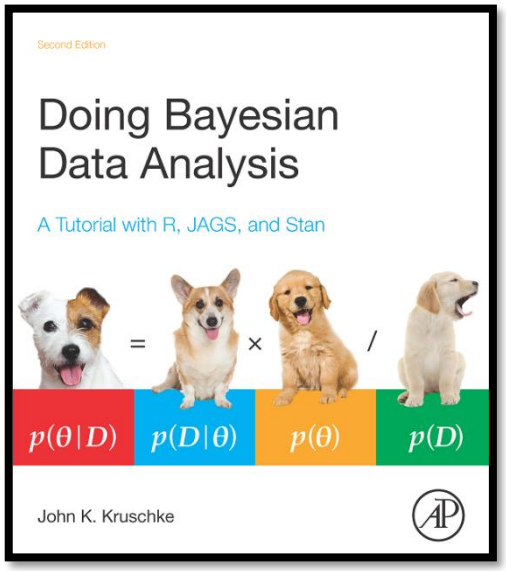
Summary

- ✓Prior
- ✓Likelihood
- ✓Posterior
- ✓Bayes' Rule
- ✓Bernoulli Distribution
- ✓Beta Distribution

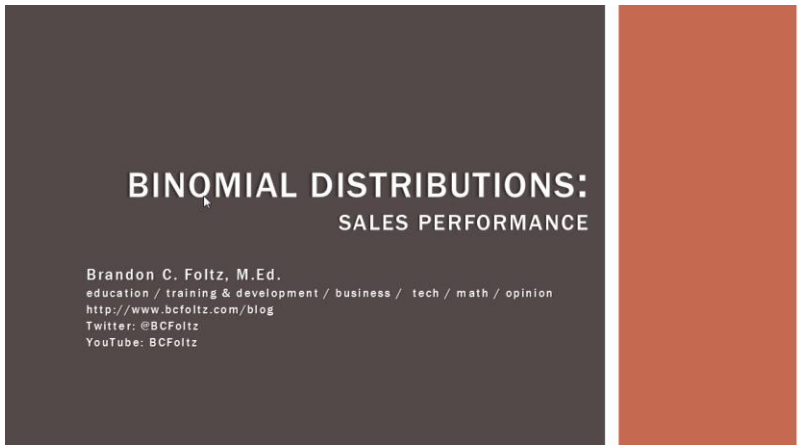


References

[Book on Doing Bayesian Data Analysis](#) - John K. Kruschke.

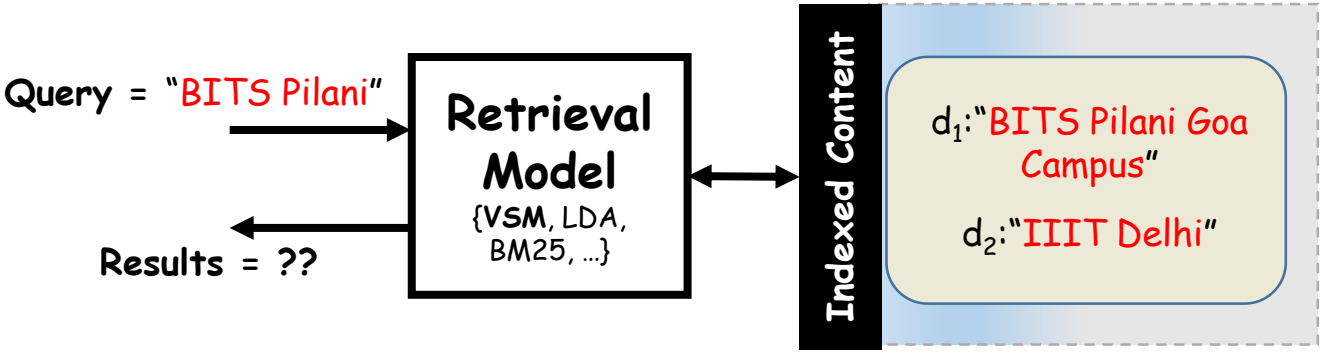


[YouTube Video on Statistics 101: The Binomial Distribution](#) - Brandon Foltz.



Designing Search Engines

Which Document to Retrieve?

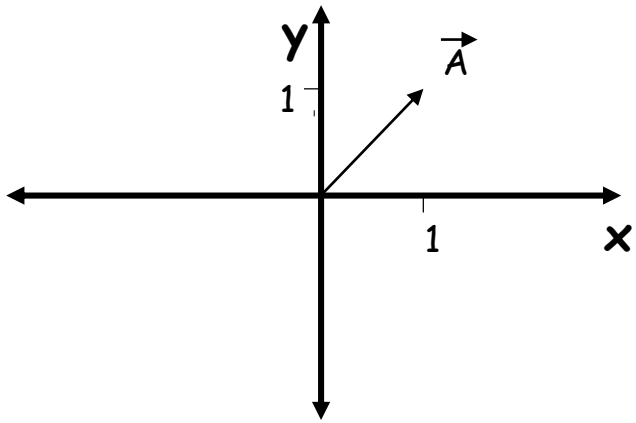


An Elegant Approach

Revisiting
Linear Algebra

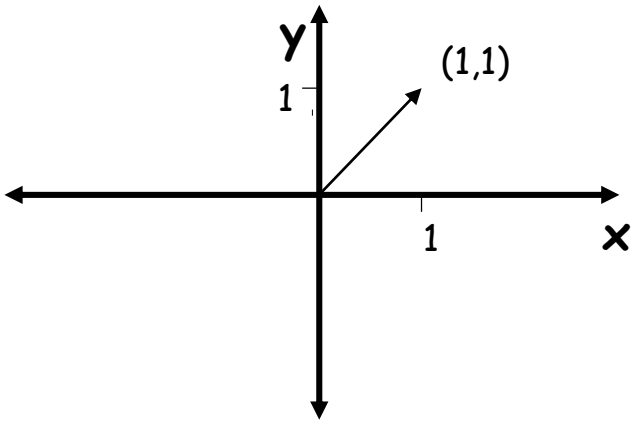
Vectors

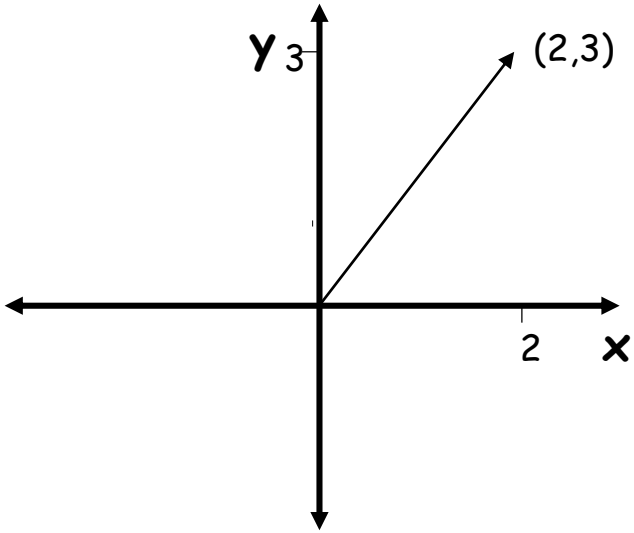
- Geometric entity which has magnitude and direction



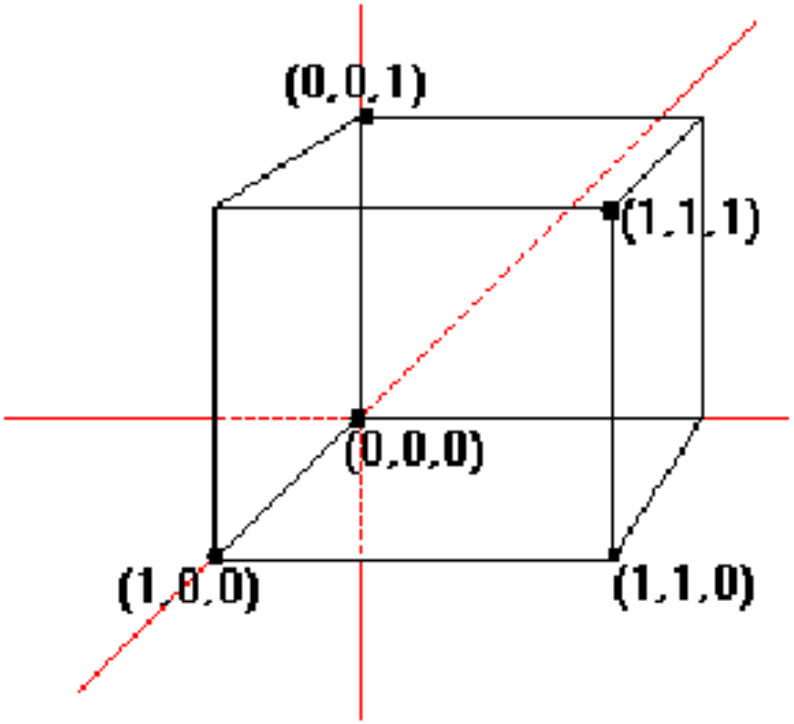
- If (x,y) is our vector of interest, this figure shows A vector = (1,1).

How is (2,3) Different?





What is (1,1,1) ?

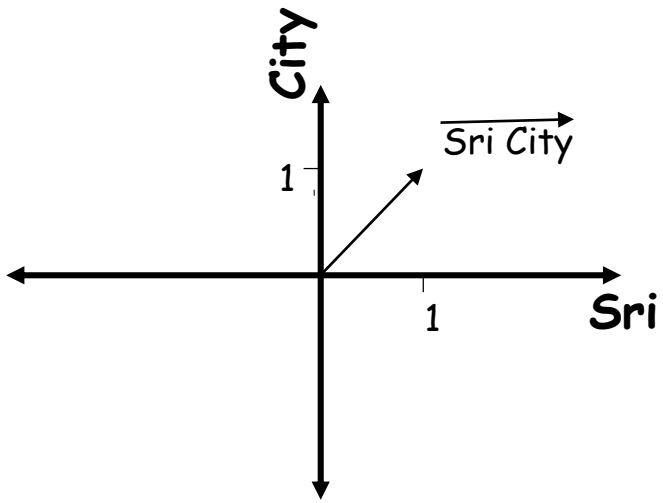


Remember!

A number is just a mathematical object. We give meaning to it!

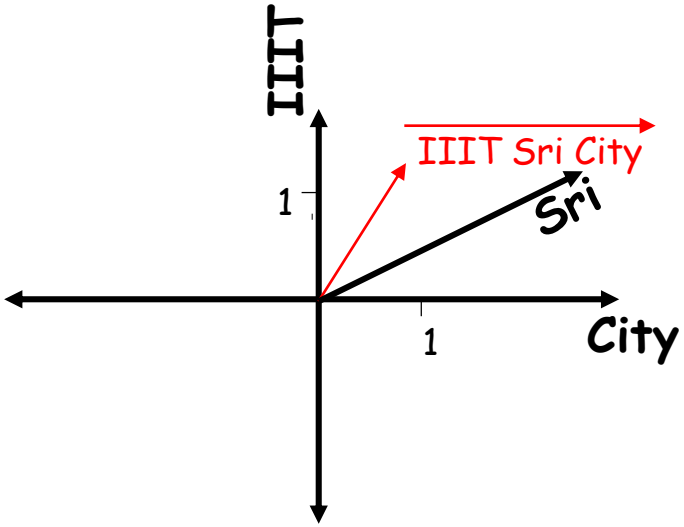
Sentences are Vectors

- "Sri City" as a vector



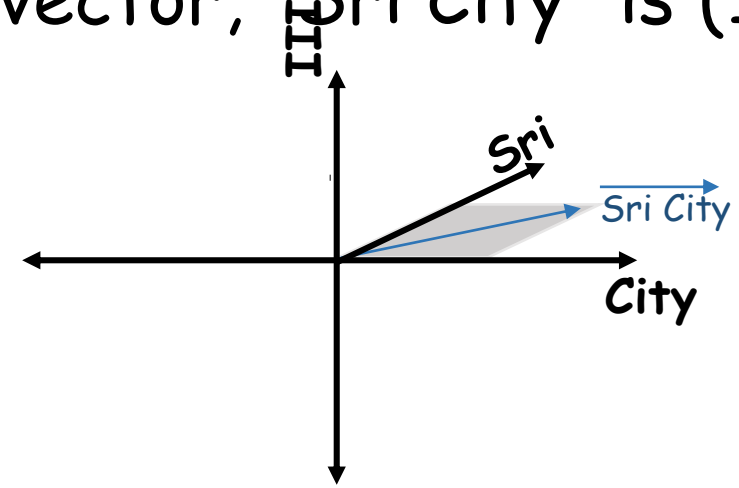
Sentences are Vectors

- "IIIT Sri City" is a 3-dimensional vector



Sentences are Vectors

- On this 3D space, "Sri City" vector will lie on the x (City) and z (Sri) plane. If (x,y,z) denotes the vector, "Sri City" is (1,0,1).



Natural Language Phrases as Vectors

Let query q = "IIIT Sri City".

Let document, d_1 = "IIIT Sri City" and d_2 = "IIIT Delhi".

	IIIT	Sri	City	Delhi
q	1	1	1	0
d_1	1	1	1	0
d_2	1	0	0	1

q = (1,1,1,0), d_1= (1,1,1,0) and d_2 = (1,0,0,1)

Quiz

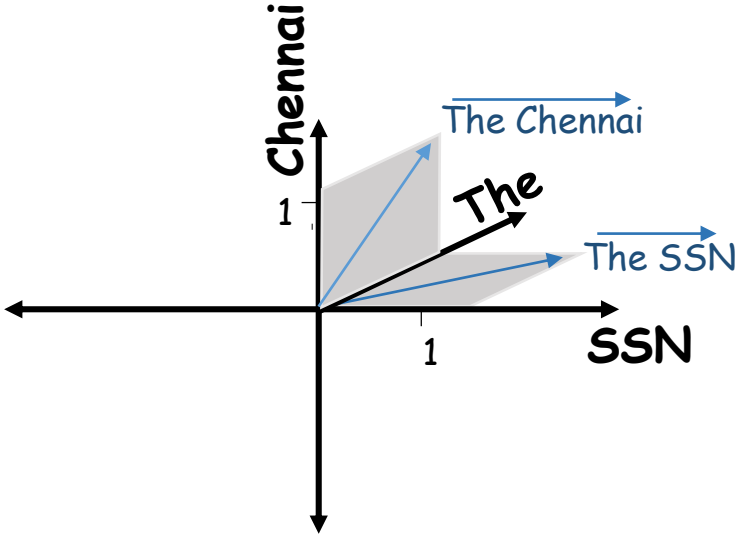
- Considering the following vectors:

	IIIT	Sri	City	Delhi
q	1	1	1	0
d_1	1	1	1	0
d_2	1	0	0	1

- What is the Natural Language (NL) equivalent of (0,1,1,0) ?
- What is the NL equivalent of (1,0,0,1) ?
- What is the vector for Delhi?
- What is the NL equivalent of q?

Comparing Sentences

- We can compare sentences using the angle between vectors



Angle between two vectors

- What is the angle between The and SSN vectors?
- What is the angle between SSN and Chennai vectors?
- What is the angle between The SSN and The SSN vectors?

Similarity Score

- D1 = "Chennai"
- D2 = "Chennai"

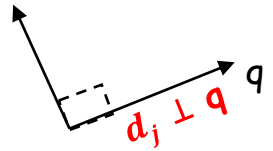
- Quiz
  - On a scale of 0 – 1, how similar are D1 and D2?
    - 0 ➔ Dissimilar
    - 1 ➔ Identical

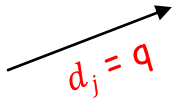
How to Convert [0 to 90] → [1 to 0]

Revisiting Trigonometry

Converting from "0 - 90" to "1 - 0"

- For convenience, We convert the angles 0 – 90 to values 1 - 0
  - When vectors are perpendicular, we want to output 0.
  - When vectors are same, we want to output 1.





0 – 90 to 1 – 0: How?

	0°	30°	45°	60°	90°
sin θ	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
cos θ	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
tan θ	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	Not defined

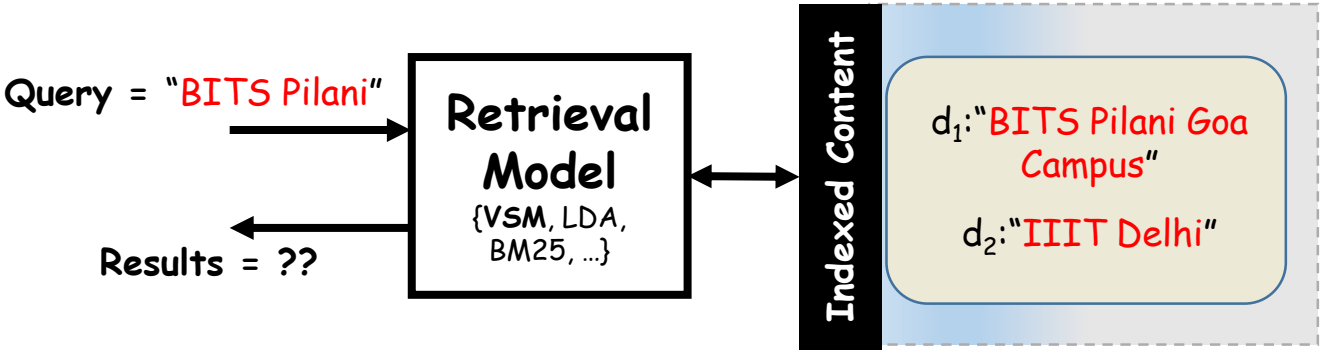
Back to Trigonometry

- If x and y are non-unit vectors, what is the cosine of angle between them (cos θ)?

$$a.b = ||a||\,||b|| \cos(\theta)$$

similarity(a,b) = $\cos(\theta) = \frac{a.b}{||a||\,||b||}$

Which Document to Retrieve?



A Boolean Term Document Matrix

	BITS	Pilani	Goa	Campus	IIIT	Delhi
q	1	1	0	0	0	0
d_1	1	1	1	1	0	0
d_2	0	0	0	0	1	1

Example

Let query q = "BITS Pilani".
Let document, d_1 = "BITS Pilani Goa Campus" and d_2 = "IIIT Delhi".

	BITS	Pilani	Goa	Campus	IIIT	Delhi
q	1	1	0	0	0	0
d_1	1	1	1	1	0	0
d_2	0	0	0	0	1	1

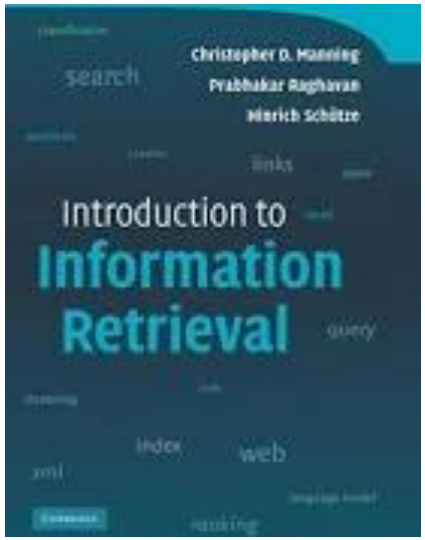
In our VSM, q = (1,1,0,0,0,0), d_1= (1,1,1,1,0,0) and d_2 = (0,0,0,0,1,1)

similarity(d_1, q) = $\frac{d_1.q}{||d_1||\,||q||} = \frac{1.1 + 1.1}{\sqrt{1^2+1^2+1^2+1^2}\,\sqrt{1^2+1^2}} = 0.71.$

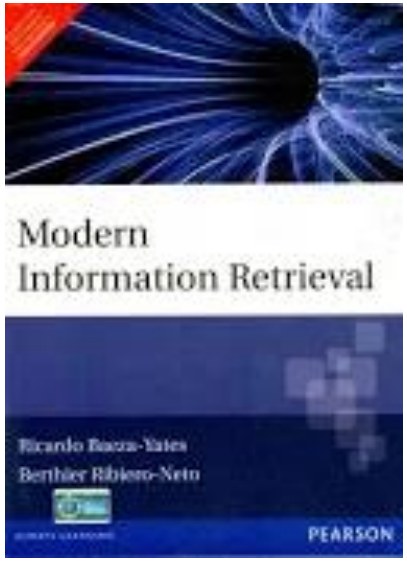
similarity(d_2, q) = $\frac{d_2.q}{||d_2||\,||q||} = 0.$

References

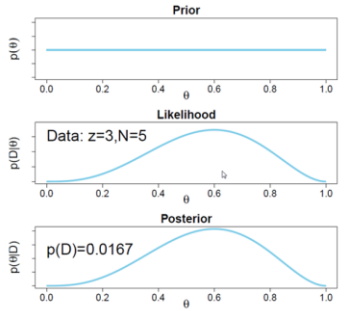
Introduction to Information Retrieval - Manning, Raghavan and Schutze.



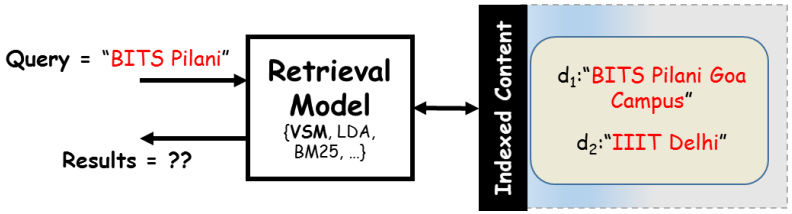
Modern Information Retrieval - Yates and Neto.



Summary



Bayesian Data Analysis



Vector Space Model for Information Retrieval

Software systems are becoming increasingly complex.

Our ability to design effective abstractions matter!

Thank You.