67-300 SEARCH ENGINES

PROBABILISTIC MODEL -BM25

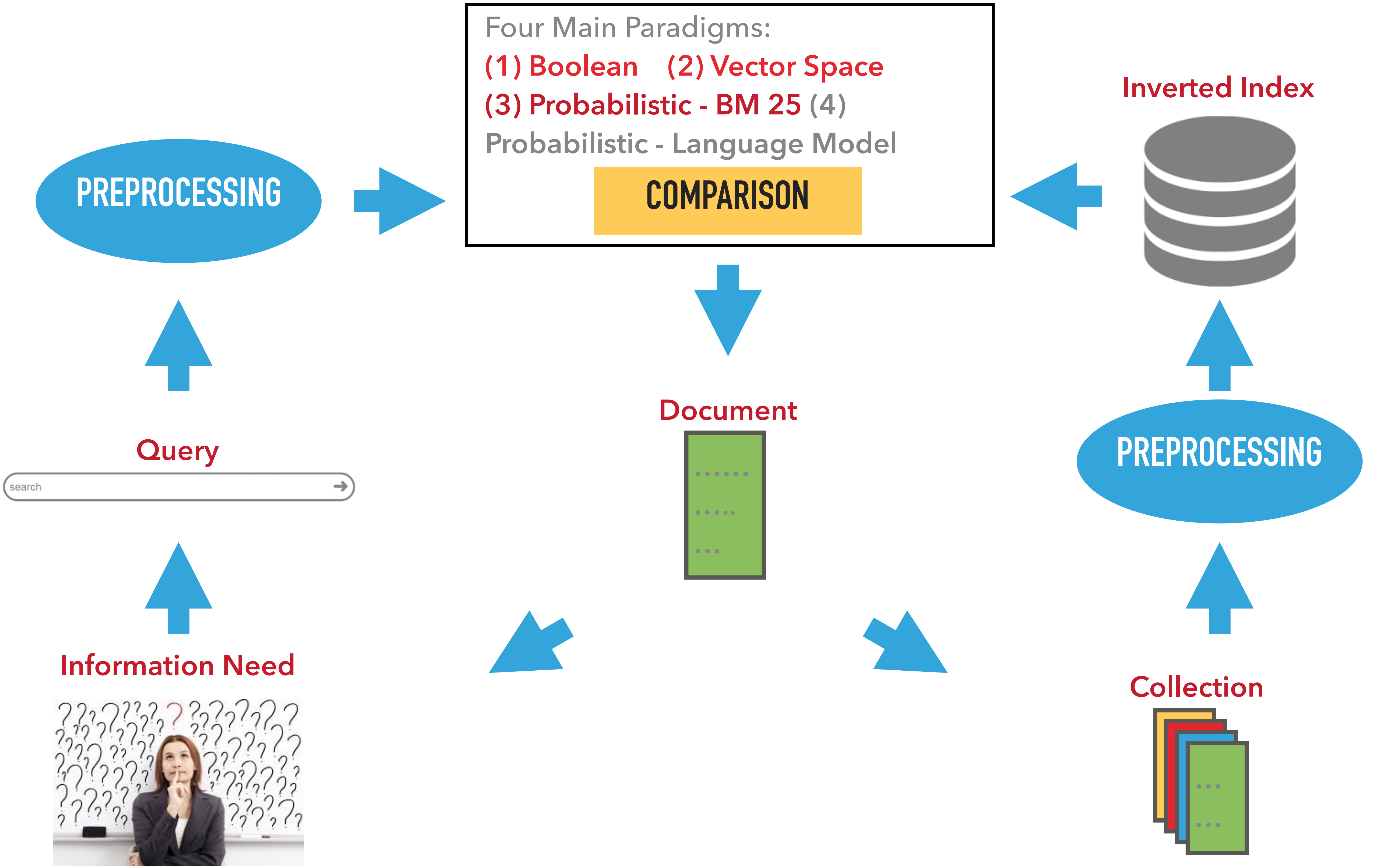
LECTURER: JOAO PALOTTI (JPALOTTI@ANDREW.CMU.EDU)

22ND MARCH 2017

LECTURE GOALS

- Notes on Normalization
- Essential concepts in probability
- Probabilistic Framework / Retrieval Status Value
- BM 25 Formula

CLASS 4 - PROBABILISTIC MODEL - BM25

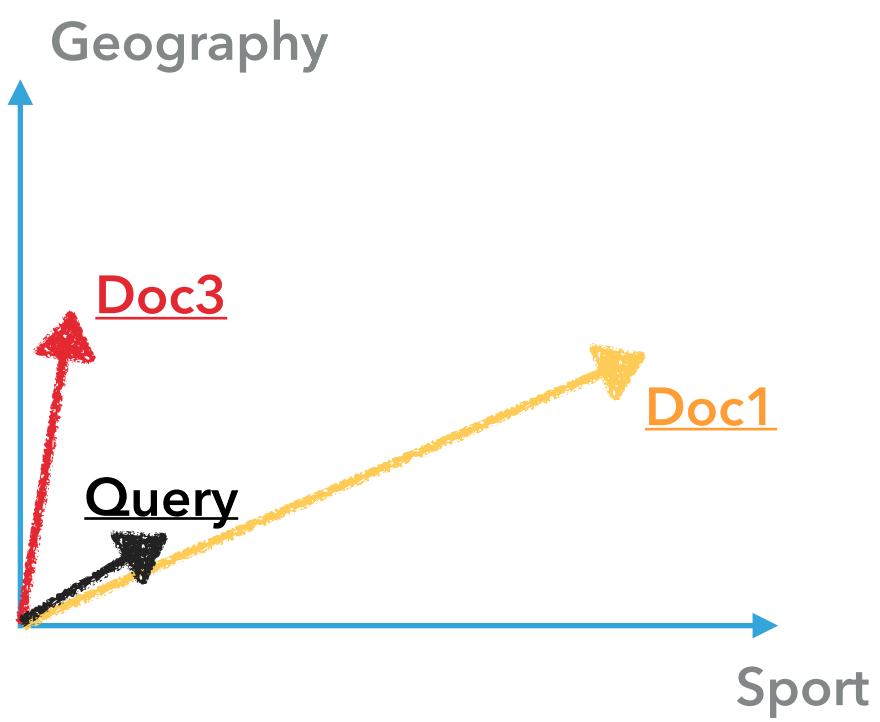


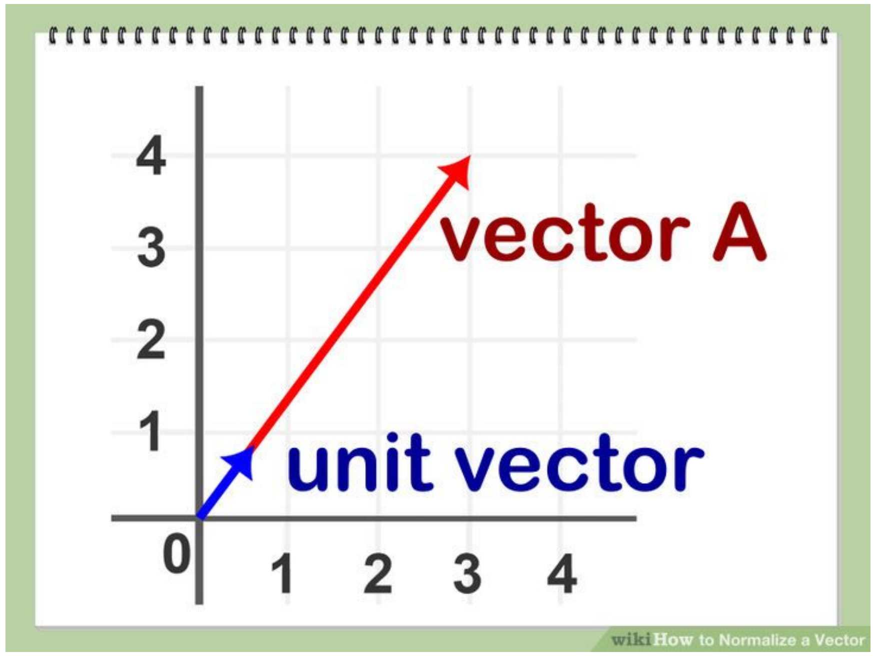
RECAP VSM

RECAP: VECTOR SPACE MODEL RECIPE

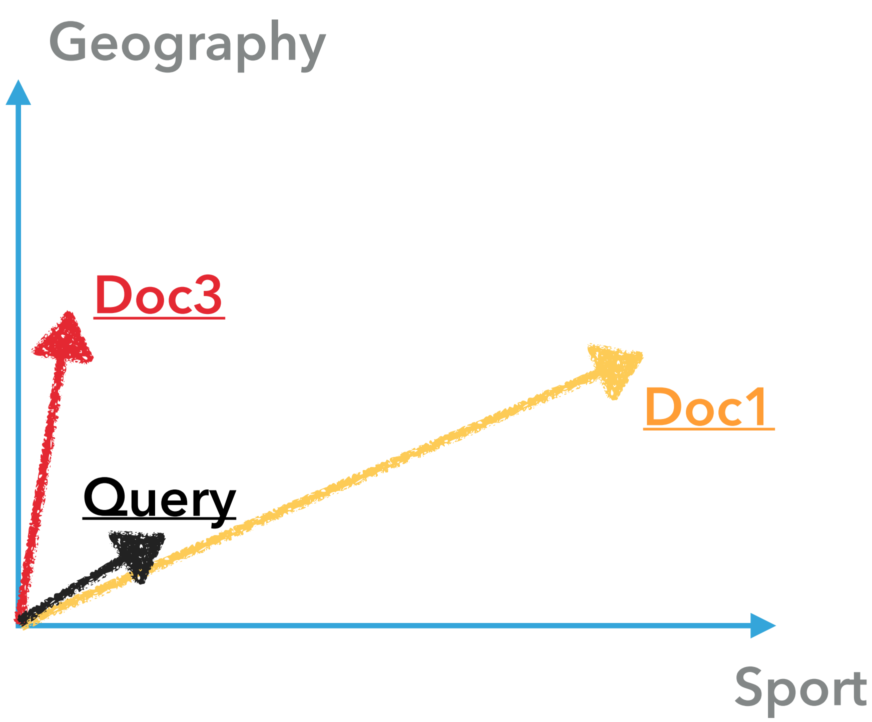
- Represent queries and documents as weighted tf-idf vectors
- Compute the cosine similarity score for the query and each document vector
- Rank documents with respect to the query by cosine similarity score
- Return top K documents (e.g., K = 10) to the user
- Geometric motivation

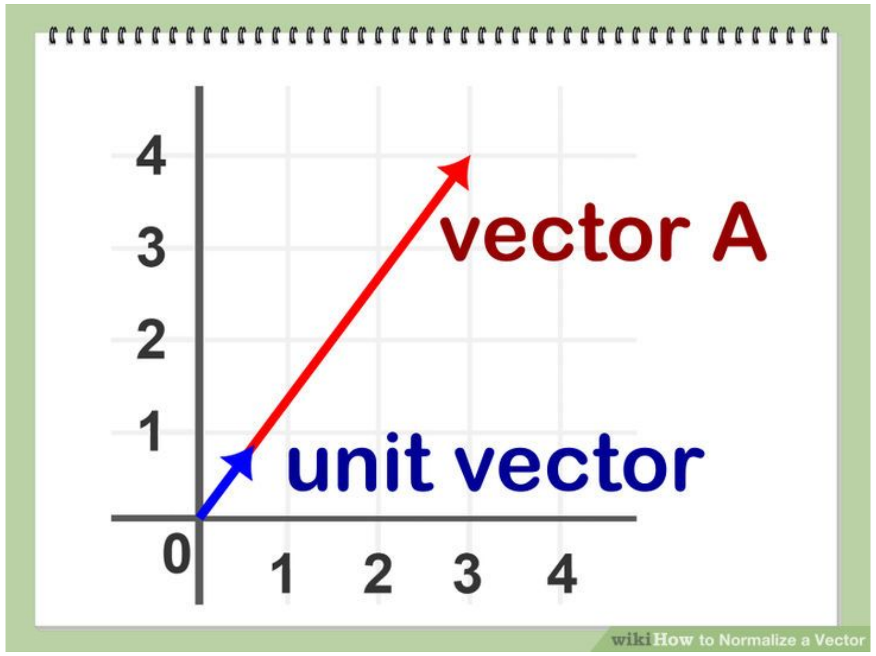
NOTE ON NORMALIZATION

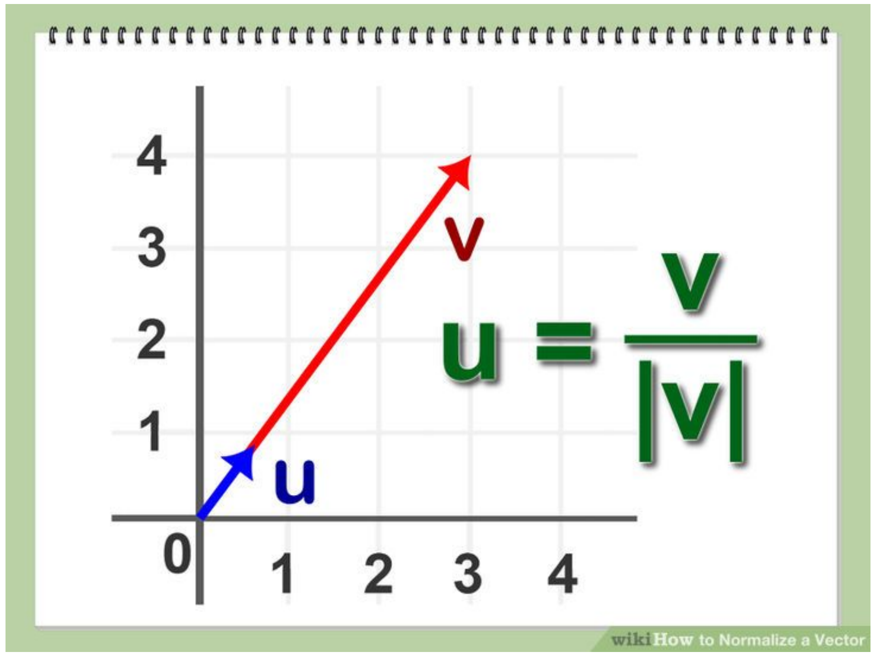




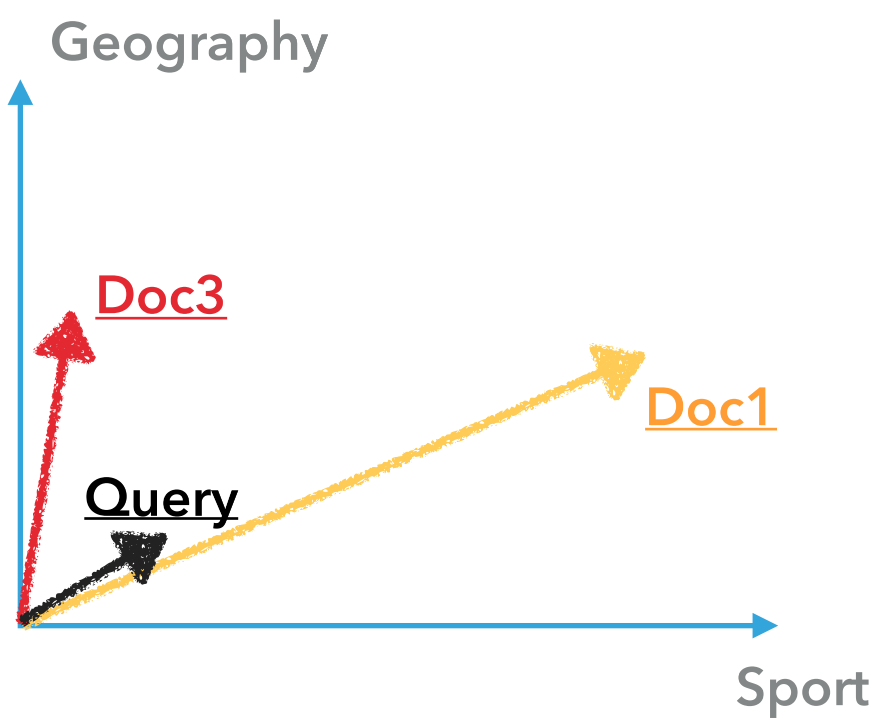
NOTE ON NORMALIZATION

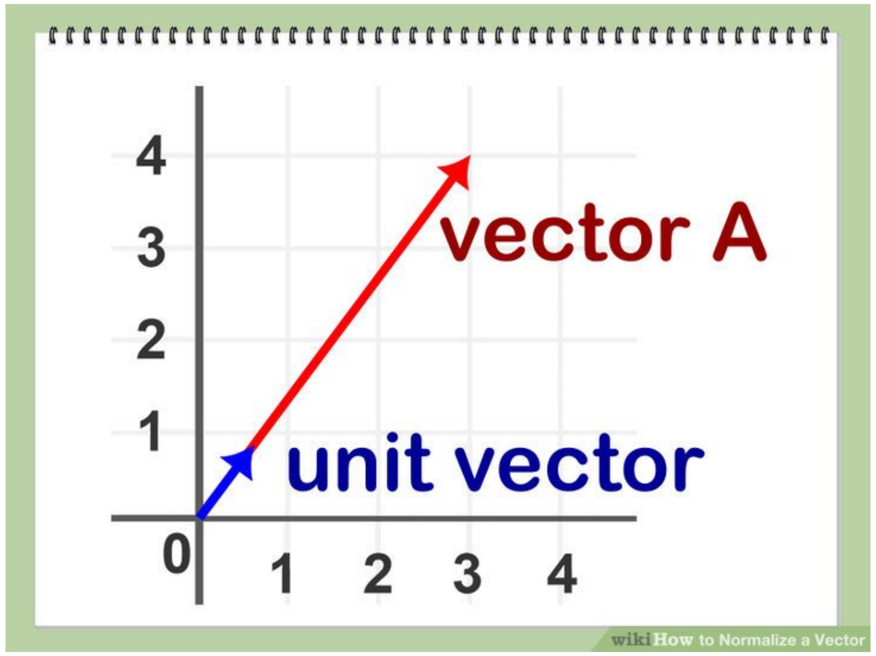


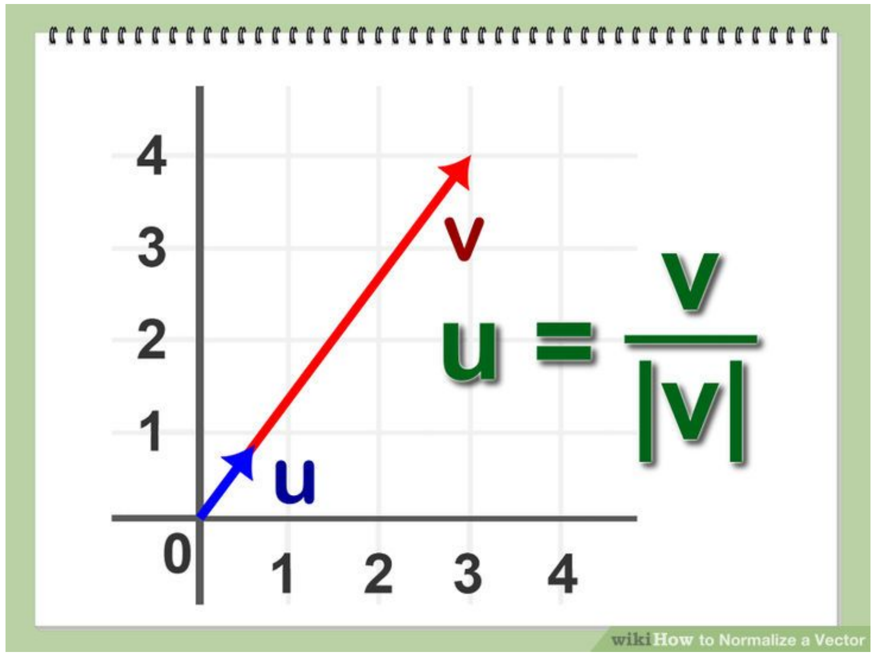


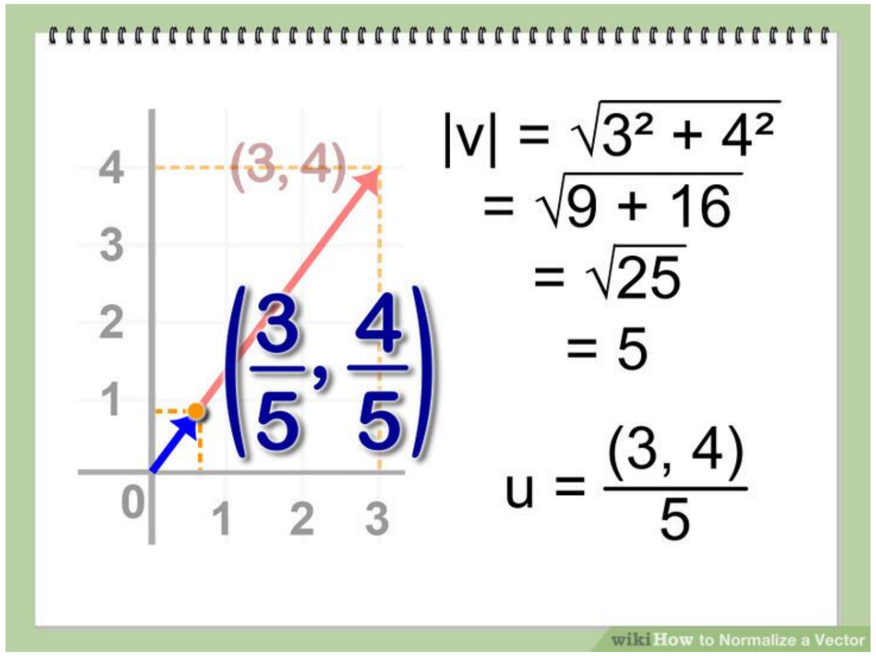


NOTE ON NORMALIZATION

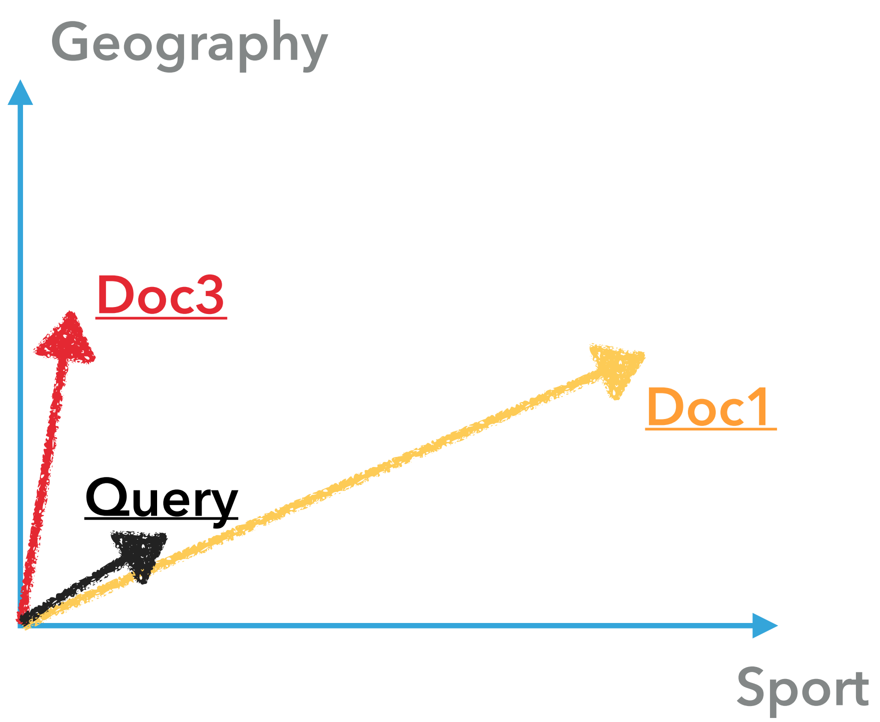




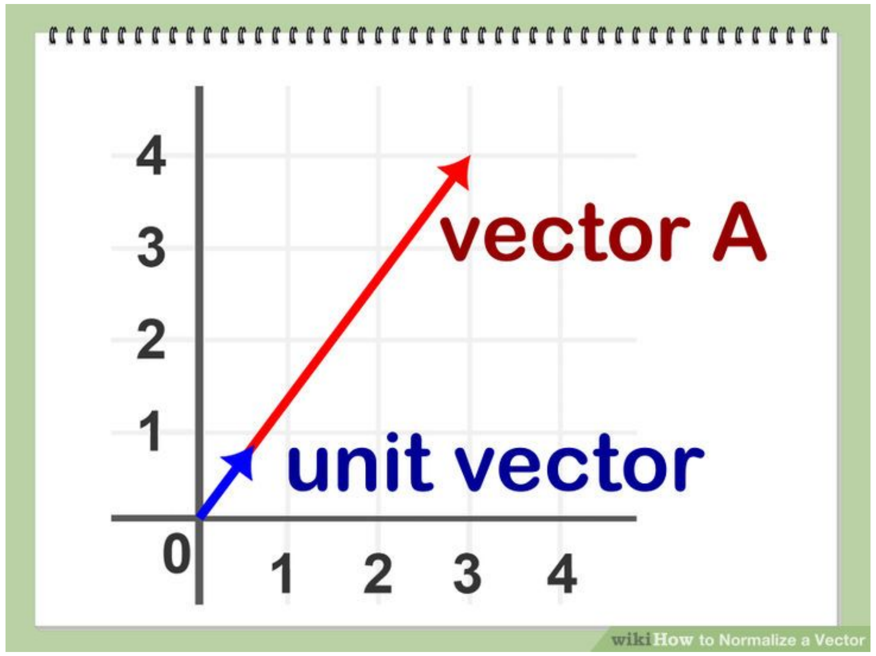


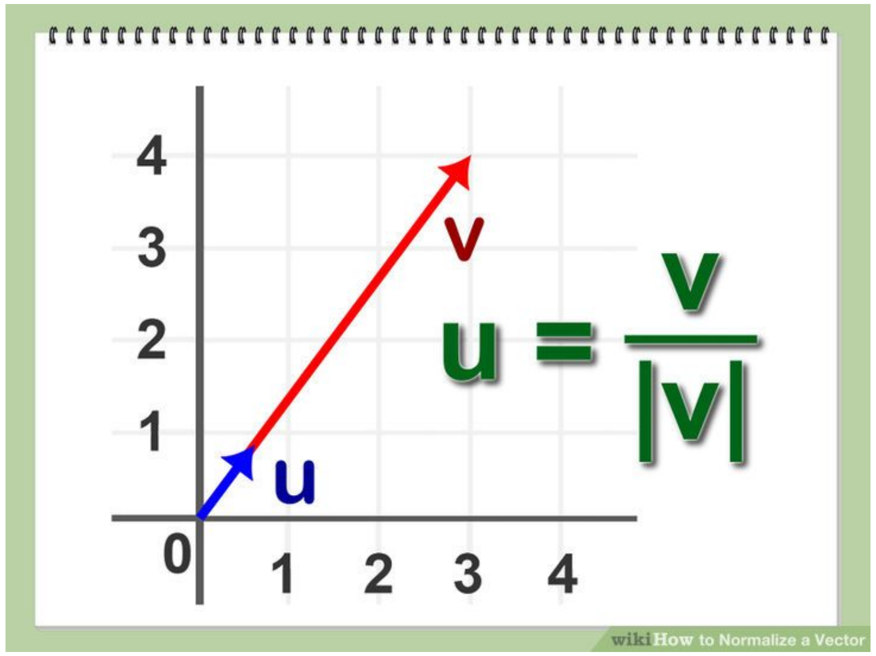


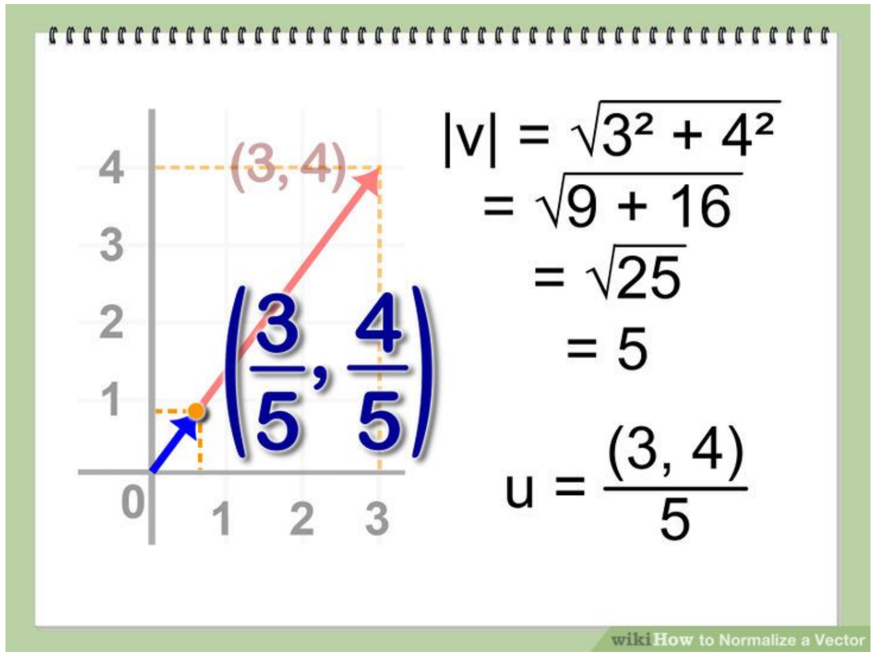
NOTE ON NORMALIZATION



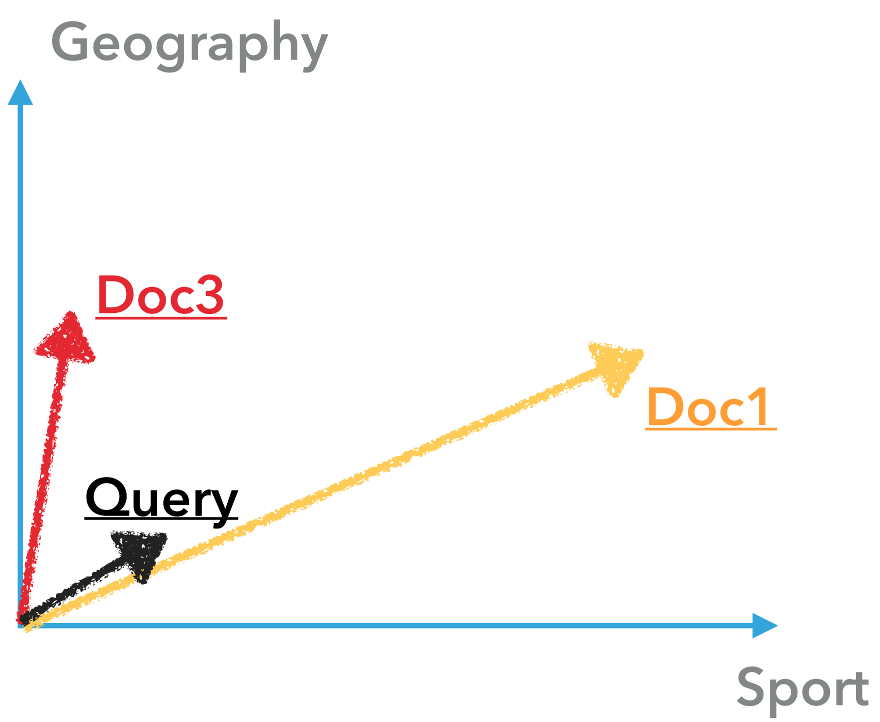
$$|u| = \sqrt{\frac{3}{5} \times \frac{3}{5} + \frac{4}{5} \times \frac{4}{5}} = 1$$

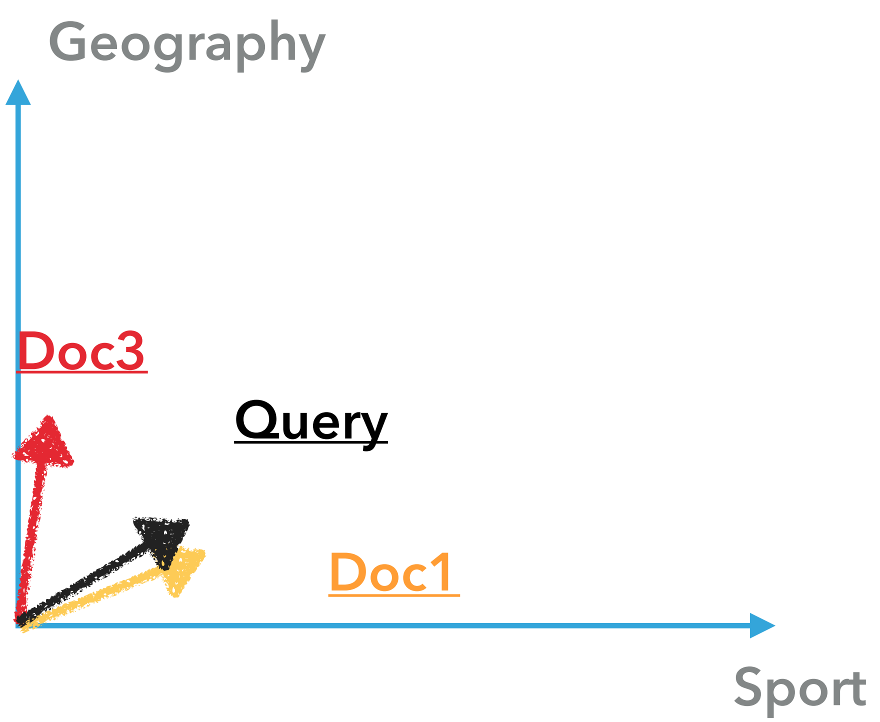


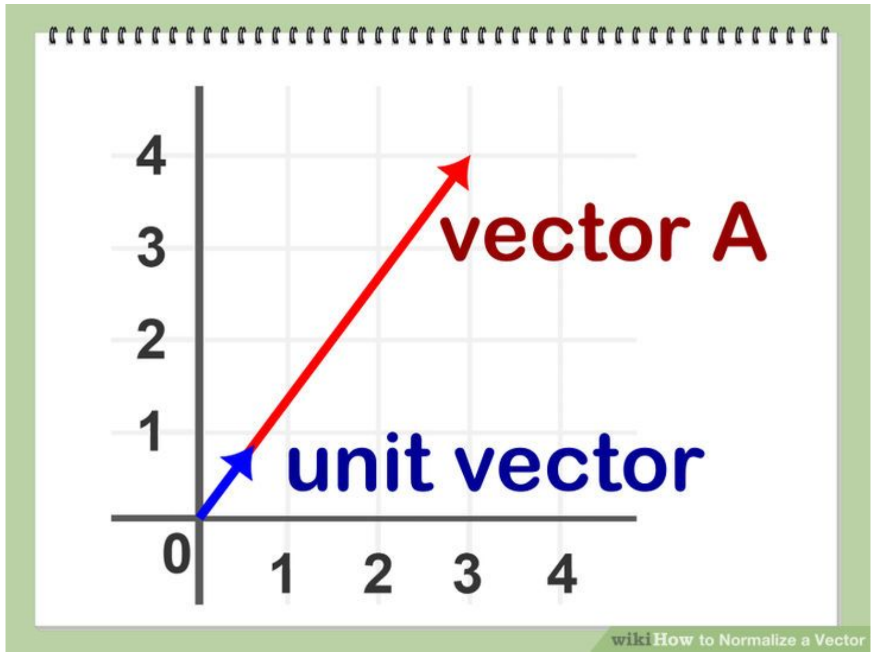


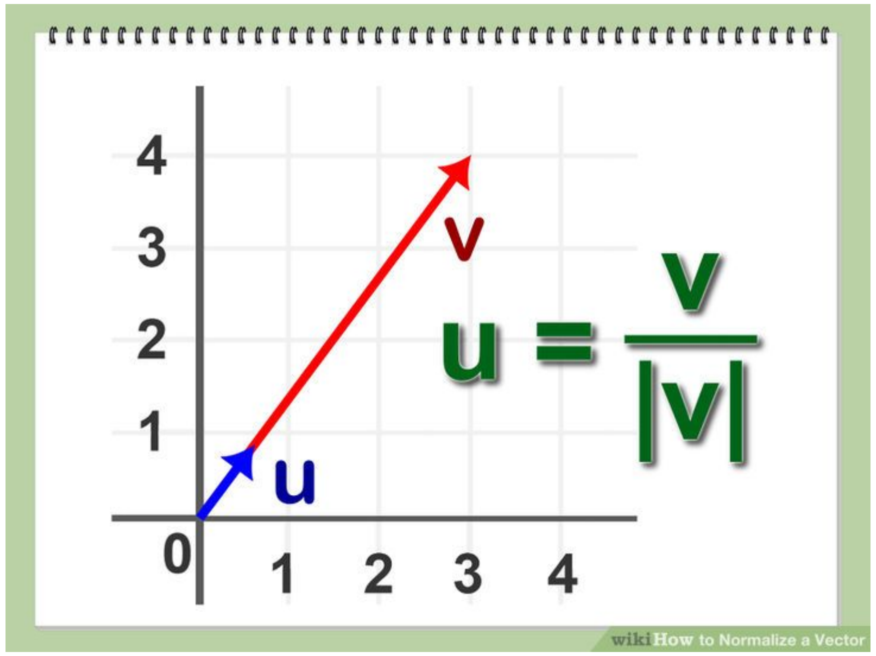


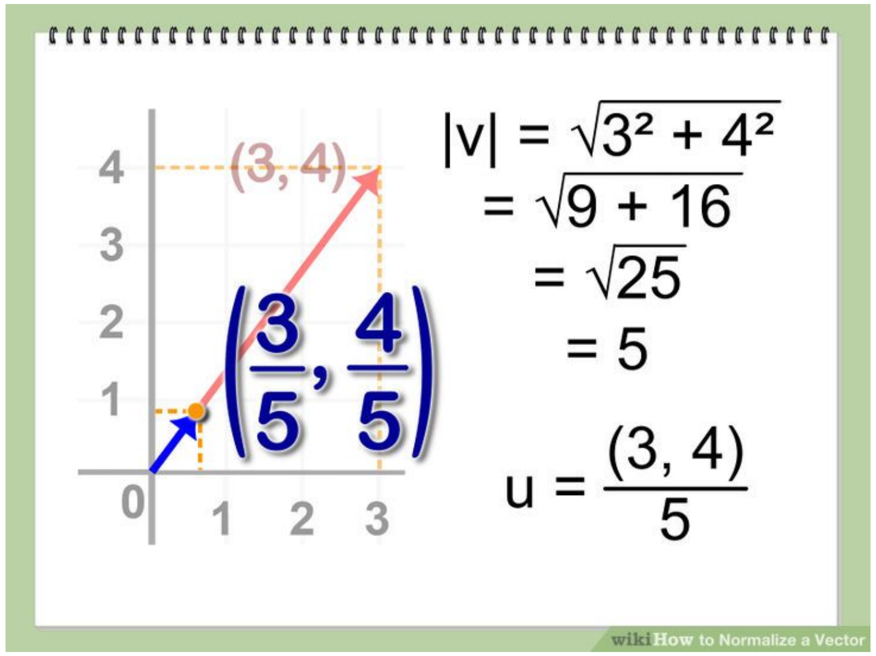
NOTE ON NORMALIZATION











GLIMPSE OF PROBABILITY

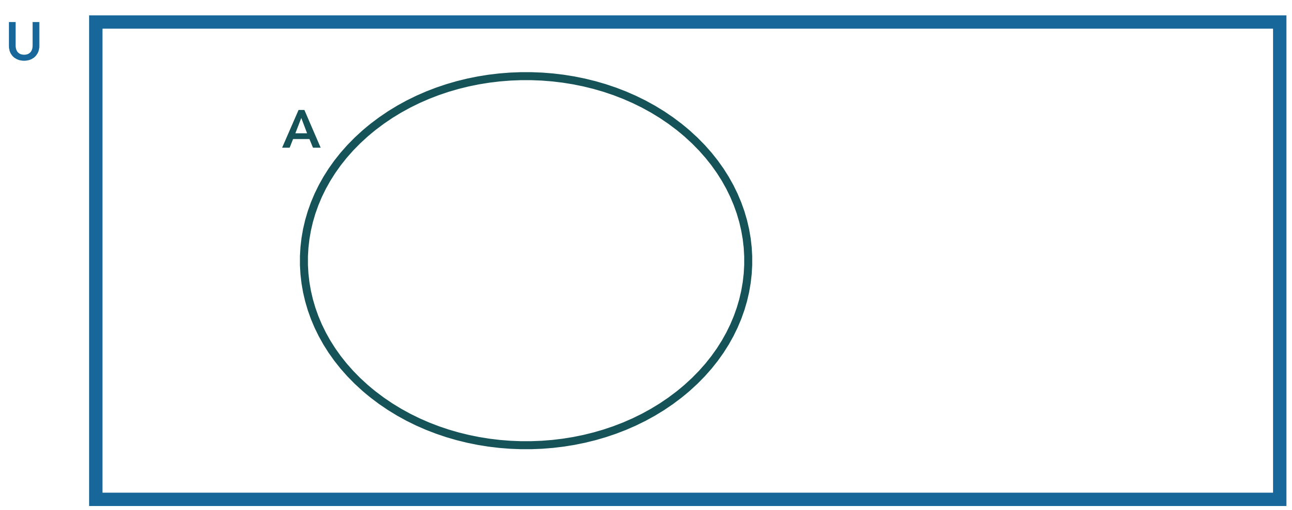
PROBABILITIES

- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$

PROBABILITIES

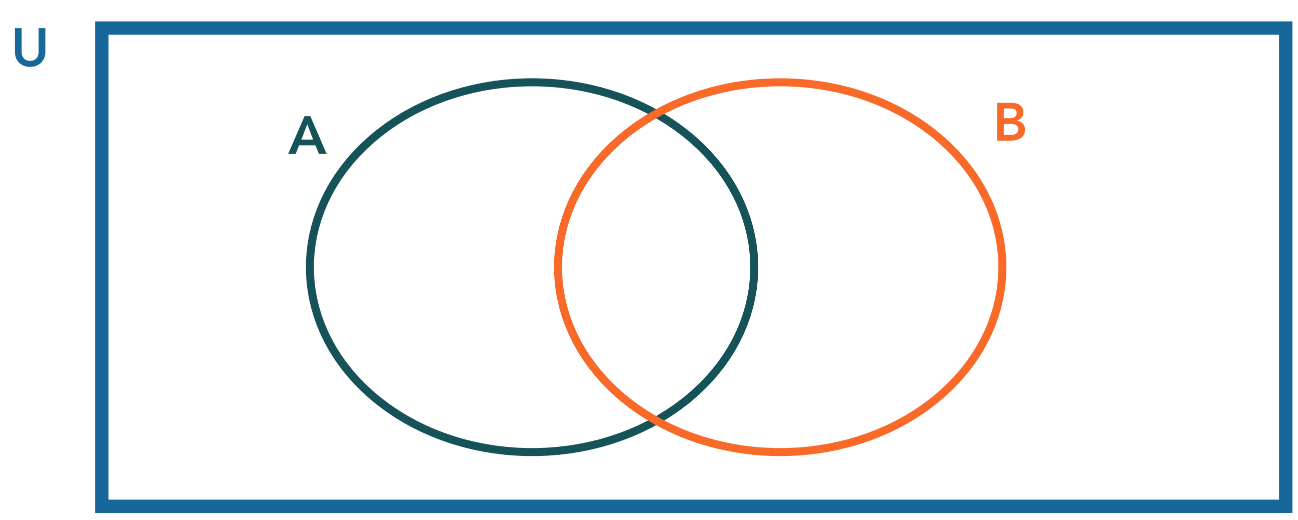
- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$

What is $P(\overline{A})$?



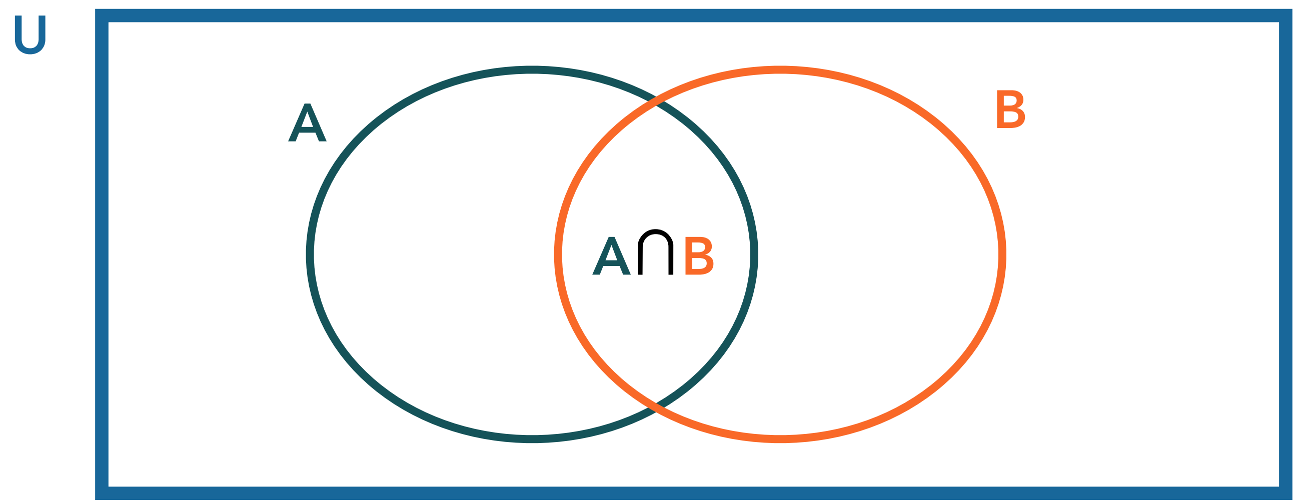
PROBABILITIES

- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$
- **Joint probability:** $P(A, B) = P(A \cap B)$



PROBABILITIES

- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$
- **Joint probability:** $P(A, B) = P(A \cap B)$



PROBABILITIES

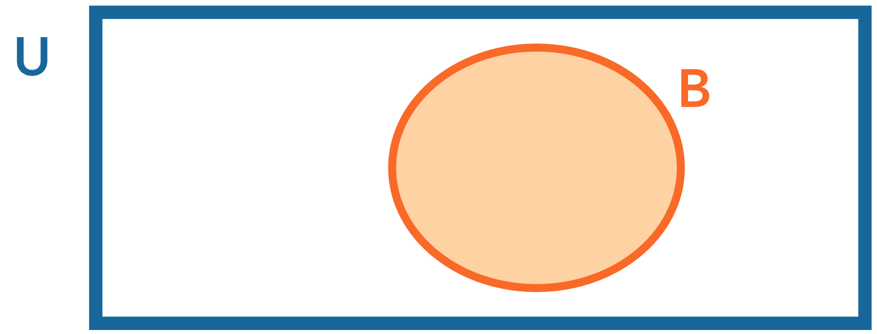
- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$
- **Joint probability:** $P(A, B) = P(A \cap B)$
- **Conditional probability:** $P(A|B)$

Probability of A given B

PROBABILITIES

- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$
- **Joint probability:** $P(A, B) = P(A \cap B)$
- **Conditional probability:** $P(A|B)$

Probability of A given B

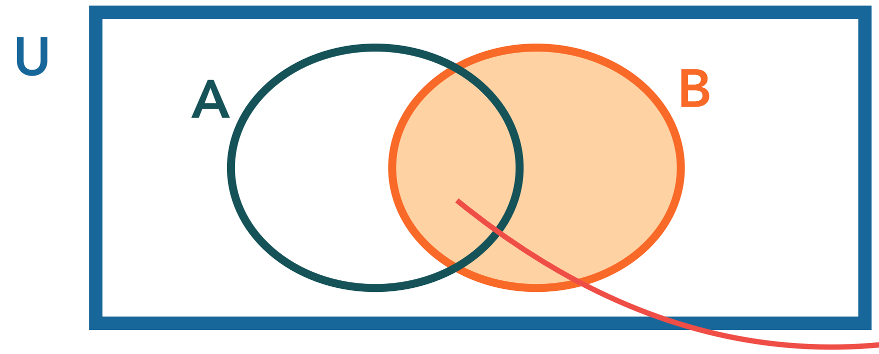


If B happened,
what is the probability of A now?

PROBABILITIES

- **Random variable** **A** – Subset of the space of possible outcomes
- $P(A)$ – Probability of event **A** happening
- $0 \leq P(A) \leq 1$
- **Joint probability:** $P(A, B) = P(A \cap B)$
- **Conditional probability:** $P(A|B)$

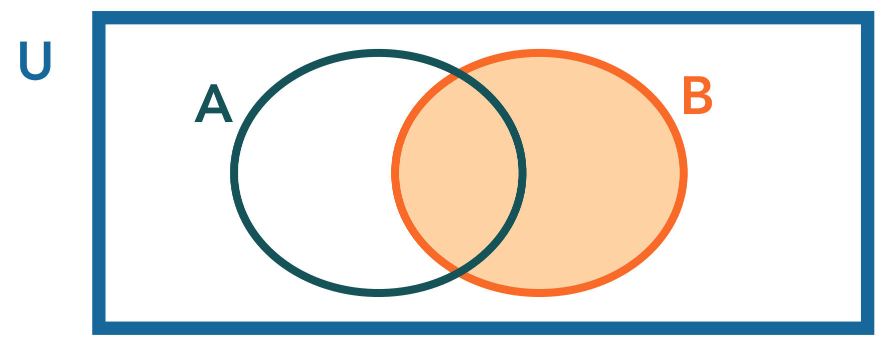
Probability of A given B



If B happened,
what is the probability of A now?

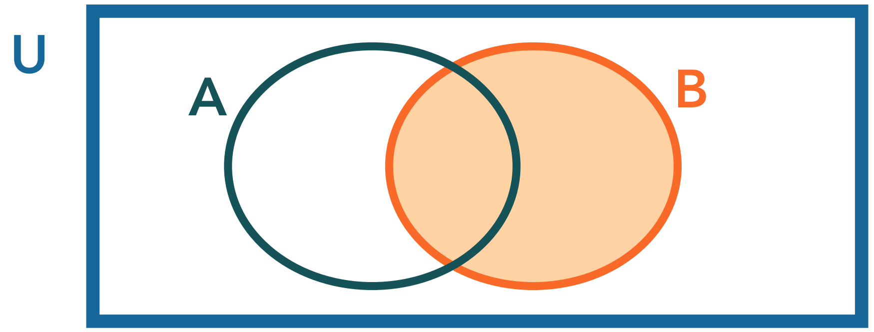
CHAIN RULE

$$P(A, B) = P(A \cap B) = P(B)P(A|B)$$

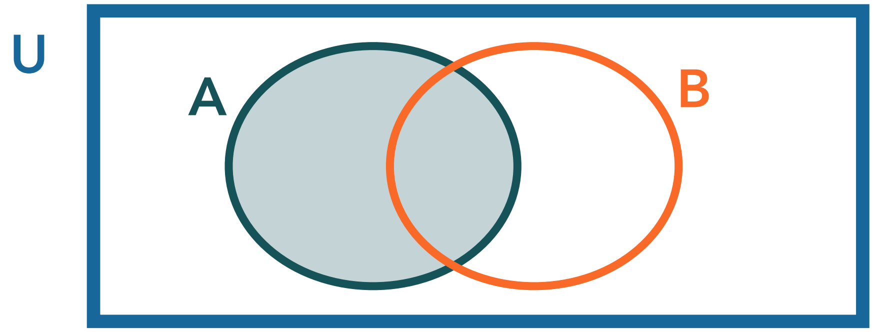


CHAIN RULE

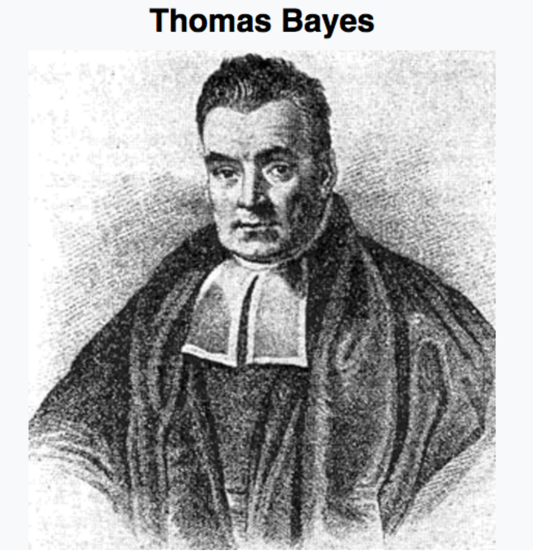
$$P(A,B) = P(A \cap B) = P(B)P(A|B)$$



$$P(A,B) = P(A \cap B) = P(B|A)P(A)$$



BAYES THEOREM



$$P(A,B) = P(A \cap B) = P(B)P(A|B)$$

$$P(A,B) = P(A \cap B) = P(B|A)P(A)$$

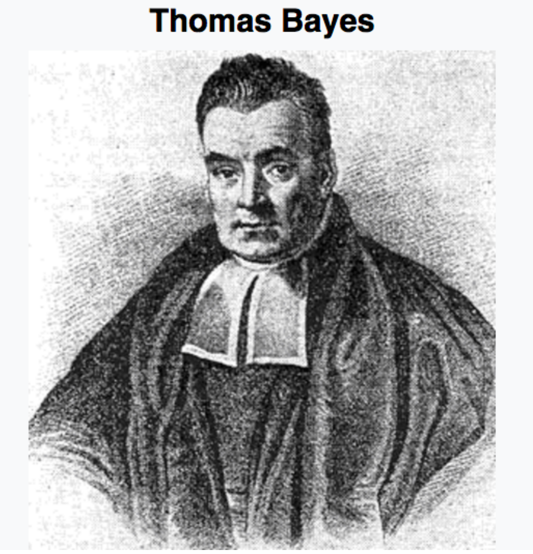
Thomas Bayes

BAYES THEOREM

$$P(A,B) = P(A \cap B) = P(B)P(A|B)$$

$$P(A,B) = P(A \cap B) = P(B|A)P(A)$$

$$P(B)P(A|B) = P(B|A)P(A)$$



BAYES THEOREM

$$P(A,B) = P(A \cap B) = P(B)P(A|B)$$

$$P(A,B) = P(A \cap B) = P(B|A)P(A)$$

$$P(B)P(A|B) = P(B|A)P(A)$$

Bayes Theorem: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$

Thomas Bayes

BAYES THEOREM

$$P(A,B) = P(A \cap B) = P(B)P(A|B)$$

$$P(A,B) = P(A \cap B) = P(B|A)P(A)$$

$$P(B)P(A|B) = P(B|A)P(A)$$

Posterior probability

Prior probability

Bayes Theorem: $$\boxed{P(A|B)} = \frac{P(B|A)\boxed{P(A)}}{P(B)}$$

OTHER SIMPLE RULES AND DEFINITIONS

- Negate one of random variables:

$$P(\overline{A}, B) = P(B|\overline{A})P(\overline{A})$$

- Interesting trivial case:

$$P(B) = P(A, B) + P(\overline{A}, B)$$

- Odds:

$$O(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1 - P(A)}$$

PROBABILISTIC FRAMEWORK

A RETRIEVAL SYSTEM RESPONSE TO A REQUEST IS A RANKING OF THE DOCUMENTS IN THE COLLECTION IN ORDER OF DECREASING PROBABILITY OF RELEVANCE TO THE USER WHO SUBMITTED THE REQUEST…

Probability Ranking Principle (PRP)

RECIPE FOR THE STATISTICAL FRAMEWORK

- Given a request **q**, for each document **d** in the collection, calculate:

$$P(R_{d,q} = 1|d, q)$$

Probability of document d being relevant for a query q, given a document d and a query q

RECIPE FOR THE STATISTICAL FRAMEWORK

- Given a request **q**, for each document **d** in the collection, calculate:

$$P(R_{d,q} = 1|d, q) \longrightarrow P(R = 1|d, q)$$

Probability of document d being relevant for a query q, given a document d and a query q

RECIPE FOR THE STATISTICAL FRAMEWORK

- Given a request **q**, for each document **d** in the collection, calculate:

$$P(R_{d,q} = 1|d, q) \longrightarrow P(R = 1|d, q)$$

- Rank documents with respect to their probability of being relevant for the query
- Return top K documents (e.g., K = 10) to the user
- Probabilistic motivation

PROBABILITY RANKING PRINCIPLE

- *x* represents a document in the collection (as a vector again)
- R represents the relevant of a document with respect to a query. R = 1, document is relevant. R = 0, document is not relevant
- $P(R = 1|x)$
- Bayes Theorem:

$$P(R = 1|x) = \frac{P(x|R = 1)P(R = 1)}{P(x)}$$

PROBABILITY RANKING PRINCIPLE

- *x* represents a document in the collection (as a vector again)
- R represents the relevant of a document with respect to a query. R = 1, document is relevant. R = 0, document is not relevant
- $P(R = 1|x)$
- Bayes Theorem:

Prior probability of retrieving a relevant document

$$P(R = 1|x) = \frac{P(x|R = 1)P(R = 1)}{P(x)}$$

probability that if a relevant document is retrieved, it is x

PROBABILITY RANKING PRINCIPLE

- *x* represents a document in the collection
- R represents the relevant of a document with respect to a query. R = 1, document is relevant. R = 0, document is not relevant
- $P(R = 1|x)$
- Bayes Theorem:

$$P(R = 1|x) = \frac{P(x|R = 1)P(R = 1)}{P(x)}$$

- $$P(R = 0|x) = \frac{P(x|R = 0)P(R = 0)}{P(x)}$$

- $P(R = 0|x) + P(R = 1|x) = 1$

PROBABILITY RANKING PRINCIPLE

- *x* represents a document in the collection
- R represents the relevant of a document with respect to a query. R = 1, document is relevant. R = 0, document is not relevant
- $P(R=1|x)$

HOW TO COMPUTE ALL THESE PROBABILITIES?
WHERE ARE THEY COMING FROM?

- Bayes Theorem:

$$P(R=1|x) = \frac{P(x|R=1)P(R=1)}{P(x)}$$

- $$P(R=0|x) = \frac{P(x|R=0)P(R=0)}{P(x)}$$

- $$P(R=0|x) + P(R=1|x) = 1$$

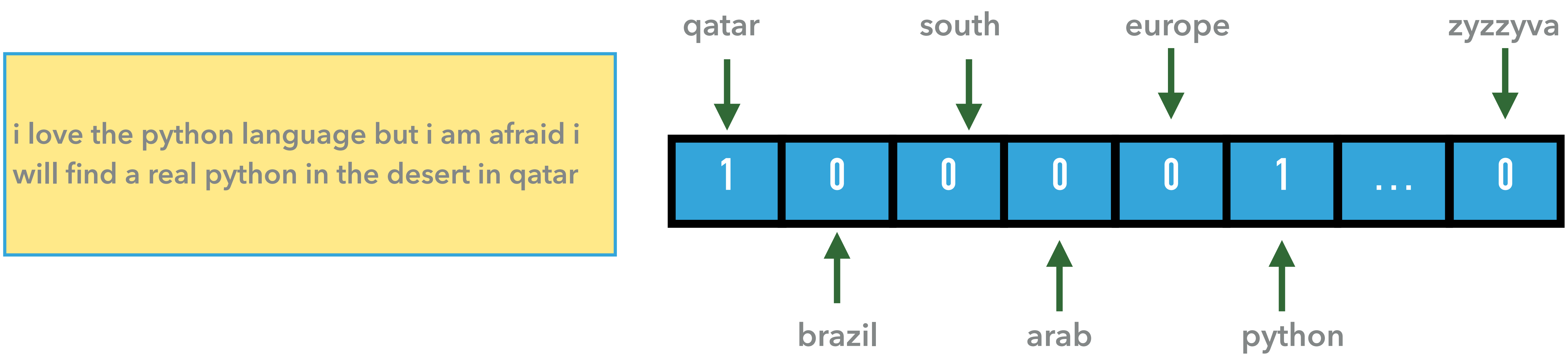
PROXIES FOR PROBABILITY

- Binary Independence Model
  - Mathematically beautiful, limited
- BM25
  - More complex theory, highly useful
- Language Models
  - A linguistic oriented approach

BINARY INDEPENDENCE MODEL

$$P(R = 1|q, x)$$

- Binary: Boolean
  - binary/Boolean version of the bag of words approach



- Independence:
  - Terms occurs in documents independently

BINARY INDEPENDENCE MODEL

$$P(R = 1|q, x)$$

- We start by calculating the odds:

$$O(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1-P(A)}$$

BINARY INDEPENDENCE MODEL

$$P(R=1|q,x)$$

- We start by calculating the odds:

$$O(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1-P(A)}$$

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)}$$

- Use Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)}$$

- Use Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\downarrow$$

$$\frac{P(R=1|q,x)}{P(R=0|q,x)} = \frac{\frac{P(R=1|q)P(x|R=1,q)}{P(x|q)}}{\frac{P(R=0|q)P(x|R=0,q)}{P(x|q)}}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)}$$

- Use Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\downarrow$$

$$\frac{P(R=1|q,x)}{P(R=0|q,x)} = \frac{\frac{P(R=1|q)P(x|R=1,q)}{\cancel{P(x|q)}}}{\frac{P(R=0|q)P(x|R=0,q)}{\cancel{P(x|q)}}}$$

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)} = \frac{P(R=1|q)}{P(R=0|q)} \times \frac{P(x|R=1,q)}{P(x|R=0,q)}$$

BINARY INDEPENDENCE MODEL

- More transformations:

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)} = \frac{P(R=1|q)}{P(R=0|q)} \times \frac{P(x|R=1,q)}{P(x|R=0,q)}$$

BINARY INDEPENDENCE MODEL

- More transformations:

Constant for a query

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)} = \boxed{\frac{P(R=1|q)}{P(R=0|q)}} \times \boxed{\frac{P(x|R=1,q)}{P(x|R=0,q)}}$$

BINARY INDEPENDENCE MODEL

- More transformations:

Constant for a query

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)} = \cancel{\boxed{\frac{P(R=1|q)}{P(R=0|q)}}} \times \boxed{\frac{P(x|R=1,q)}{P(x|R=0,q)}}$$

BINARY INDEPENDENCE MODEL

- More transformations:

Constant for a query

$$O(R|q,x) = \frac{P(R=1|q,x)}{P(R=0|q,x)} = \cancel{\boxed{\frac{P(R=1|q)}{P(R=0|q)}}} \times \boxed{\frac{P(x|R=1,q)}{P(x|R=0,q)}}$$

Still needs estimation

- Use independence assumption:

$$\frac{P(\overrightarrow{x}|R=1,q)}{P(\overrightarrow{x}|R=0,q)} = \prod_{i=1}^{V} \frac{P(x_i|R=1,q)}{P(x_i|R=0,q)}$$

BINARY INDEPENDENCE MODEL

- That's where we are so far: $O(R|q,x) = \prod_{i=1}^{V} \frac{P(x_i|R=1,q)}{P(x_i|R=0,q)}$

BINARY INDEPENDENCE MODEL

- That's where we are so far: $O(R|q,x) = \prod_{i=1}^{V} \frac{P(x_i|R=1,q)}{P(x_i|R=0,q)}$
- We can divide it into two:

$$O(R|q,x) = \prod_{x_i=1} \frac{P(x_i=1|R=1,q)}{P(x_i=1|R=0,q)} \prod_{x_i=0} \frac{P(x_i=0|R=1,q)}{P(x_i=0|R=0,q)}$$

BINARY INDEPENDENCE MODEL

- That's where we are so far: $O(R|q,x) = \prod_{i=1}^{V} \frac{P(x_i|R=1,q)}{P(x_i|R=0,q)}$
- We can divide it into two:

$$O(R|q,x) = \prod_{x_i=1} \frac{P(x_i=1|R=1,q)}{P(x_i=1|R=0,q)} \prod_{x_i=0} \frac{P(x_i=0|R=1,q)}{P(x_i=0|R=0,q)}$$

- p_i : term appearing in a relevant document

$$p_i = P(x_i=1|R=1,q)$$

- r_i : term appearing in a non relevant document

$$r_i = P(x_i=1|R=0,q)$$

BINARY INDEPENDENCE MODEL

- We can rewrite from:

$$O(R|q,x) = \prod_{x_i=1} \frac{P(x_i=1|R=1,q)}{P(x_i=1|R=0,q)} \prod_{x_i=0} \frac{P(x_i=0|R=1,q)}{P(x_i=0|R=0,q)}$$

$$p_i = P(x_i=1|R=1,q) \qquad r_i = P(x_i=1|R=0,q)$$

- To:

$$O(R|q,x) = \prod_{x_i=1} \frac{p_i}{r_i} \prod_{x_i=0} \frac{1-p_i}{1-r_i}$$

BINARY INDEPENDENCE MODEL

- Not done yet…

$$O(R|q,x) = \prod_{x_i=1} \frac{p_i}{r_i} \prod_{x_i=0} \frac{1-p_i}{1-r_i}$$

- We assume terms not in the query can be ignored:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- There is still another trick to manipulate this equation…

BINARY INDEPENDENCE MODEL

- What we have: $$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

BINARY INDEPENDENCE MODEL

- What we have: $$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- After adding a small trick here:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=1;q_i=1} \frac{1-r_i}{1-p_i} \times \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

BINARY INDEPENDENCE MODEL

- What we have: $$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- After adding a small trick here:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=1;q_i=1} \frac{1-r_i}{1-p_i} \times \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- And moving things around:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} \prod_{x_i=1;q_i=1} \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

BINARY INDEPENDENCE MODEL

- What we have: $$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- After adding a small trick here:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=1;q_i=1} \frac{1-r_i}{1-p_i} \times \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- And moving things around:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} \boxed{\prod_{x_i=1;q_i=1} \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}}$$

BINARY INDEPENDENCE MODEL

- What we have:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- After adding a small trick here:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i}{r_i} \prod_{x_i=1;q_i=1} \frac{1-r_i}{1-p_i} \times \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}$$

- And moving things around:

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} \boxed{\cancel{\prod_{x_i=1;q_i=1} \frac{1-p_i}{1-r_i} \prod_{x_i=0;q_i=1} \frac{1-p_i}{1-r_i}}}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

- Multiplying fractions is hard for computers, we can use the logarithm to deal with hardware/numeric limitation
- Result is know as **Retrieval Status Value (RSV)**

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

- Multiplying fractions is hard for computers, we can use the logarithm to deal with hardware/numeric limitation
- Result is know as **Retrieval Status Value (RSV)**

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \underbrace{\boxed{\frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}}}_{c_i}$$

BINARY INDEPENDENCE MODEL

$$O(R|q,x) = \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

- Multiplying fractions is hard for computers, we can use the logarithm to deal with hardware/numeric limitation
- Result is know as **Retrieval Status Value (RSV)**

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \boxed{\frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}} \quad c_i$$

How can we compute the cs (ps and rs)?

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$			n
$x_i=0$			
Total			N

df

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$			n
$x_i=0$			N - n
Total			N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$			n
$x_i=0$			N - n
Total	S		N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$	s		n
$x_i=0$			N - n
Total	S		N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$	s		n
$x_i=0$	S-s		N - n
Total	S		N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
x_i=1	s		n
x_i=0	S-s		N - n
Total	S	N-S	N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$	s	n-s	n
$x_i=0$	S-s		N - n
Total	S	N-S	N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i*:

Docs	Relevant	Non-Rel.	Total
$x_i=1$	s	n-s	n
$x_i=0$	S-s	N-n-S+s	N - n
Total	S	N-S	N

ESTIMATING RSV COEFFICIENTS IN THEORY

- Table for each term *i:*

Docs	Relevant	Non-Rel.	Total
$x_i=1$	s	n-s	n
$x_i=0$	S-s	N-n-S+s	N - n
Total	S	N-S	N

- $$p_i \approx \frac{s}{S}$$
- $$r_i \approx \frac{n-s}{N-S}$$

ESTIMATING RSV COEFFICIENTS IN THEORY

- Merging everything into…

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

$$p_i \approx \frac{s}{S} \qquad r_i \approx \frac{n-s}{N-S}$$

$$c_i = K(N,n,S,s) = \log \frac{\frac{s}{S-s}}{\frac{n-s}{N-n-S+s}}$$

KEY RESULTS FROM THE THEORY

- Given the RSV value, what if p ~ 0?

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

$$RSV = \log \sum_{x_i=1;q_i=1} \frac{(1-r_i)}{r_i} \quad \longrightarrow \quad \frac{N-n-S+s}{n-s}$$

KEY RESULTS FROM THE THEORY

- Given the RSV value, what if p ~ 0?

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

$$RSV = \log \sum_{x_i=1;q_i=1} \frac{(1-r_i)}{r_i} \dashrightarrow \frac{N-n-S+s}{n-s}$$

- If p ~ 0, then s ~ 0. Results in: $\log \frac{N-n}{n} \approx \log \frac{N}{n}$

KEY RESULTS FROM THE THEORY

- Given the RSV value, what if p ~ 0?

$$RSV = \log \prod_{x_i=1;q_i=1} \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)} = \sum_{x_i=1;q_i=1} \log \frac{p_i \times (1-r_i)}{r_i \times (1-p_i)}$$

$$RSV = \log \sum_{x_i=1;q_i=1} \frac{(1-r_i)}{r_i} \quad \cdots\cdots\cdots\rightarrow \quad \frac{N-n-S+s}{n-s}$$

- If p ~ 0, then s ~ 0. Results in: $\log \frac{N-n}{n} \approx \log \frac{N}{n}$ ➡ <u>IDF!!!</u>

ESTIMATION OF P IS THE HARDEST PART

- Remember p_i is the probability of term i in relevant documents
- Getting an accurate estimation of p_i is hard (but not impossible)
- Proxies:
  - From a set of known relevant documents (pseudo-relevance)
  - A constant value – Then we use only IDF
  - proportional to the prob. of occurrence of term i in the collection

BOOTSTRAPPING P_I

1. Assume p_i is a constant (rank using only IDF)
2. Ask the user/Guess the relevant document set **D**
3. Improve estimation for p_i and r_i
    1. Count distribution of x_i in D. Adjust $p_i = |D_i| / |D|$
    2. Not retrieved documents counted as not relevant. Adjust $r_i = (n_i - |D_i|) / (N - |D|)$
4. Repeat from 2 until p_i and r_i converge.

BM25

BM 25

- Best Match 25 results in a series of empirical try-and-error
- Aims to overcome some limitations from BIM, such as the binary part

$$RSV_d = \sum_{t \in q} \log \left[\frac{N}{df_t} \right] \cdot \frac{(k_1 + 1) tf_{td}}{k_1((1-b) + b \times (L_d / L_{ave})) + tf_{td}}$$

BM 25

- Best Match 25 results in a series of empirical try-and-error
- Aims to overcome some limitations from BIM, such as the binary part

$$RSV_d = \sum_{t \in q} \boxed{\log \left[\frac{N}{df_t} \right]} \cdot \boxed{\frac{(k_1 + 1) tf_{td}}{k_1((1 - b) + b \times (L_d / L_{ave})) + tf_{td}}}$$

BM 25

- Best Match 25 results in a series of empirical try-and-error
- Aims to overcome some limitations from BIM, such as the binary part

$$RSV_d = \sum_{t \in q} \boxed{\log\left[\frac{N}{df_t}\right]} \cdot \boxed{\frac{(k_1 + 1)tf_{td}}{k_1((1-b) + b \times (L_d/L_{ave})) + tf_{td}}}$$

<u>**What if we have b = 0?**</u>

<u>**What if we have b = 1?**</u>

BM 25

- Best Match 25 results in a series of empirical try-and-error
- Aims to overcome some limitations from BIM, such as the binary part

$$RSV_d = \sum_{t \in q} \boxed{\log\left[\frac{N}{df_t}\right]} \cdot \boxed{\frac{(k_1 + 1)tf_{td}}{k_1((1-b) + b \times (L_d/L_{ave})) + tf_{td}}}$$

What if we have b = 0?

What if we have b = 1?

What if we have K1 = 0?

What if we have K1 very high?

BM 25

- We might have large queries and we might want to control for query size as well:

$$RSV_d = \sum_{t \in q} \boxed{\log \left[\frac{N}{df_t} \right]} \cdot \boxed{\frac{(k_1 + 1) tf_{td}}{k_1((1-b) + b \times (L_d / L_{ave})) + tf_{td}}} \boxed{\frac{(k_3 + 1) tf_{tq}}{k_3 + tf_{tq}}}$$

- Typical parameters are:
  - 1.2 < K_1 < 2 ; 0 < K_3 < 1000; b = 0.75
- It is common to use the smoothed version of BM25. We will see what is smoothing next lecture…

WHAT DID WE SEE? WHAT SHOULD YOU KNOW?

- Essential concepts in probability
- Theoretic justification of ranking by relevance
- Derivation of the Retrieval Status Value (RSV)
- BM 25

TODAY'S LECTURE IN THE STANFORD IR BOOK

- Chapter 11 - Probabilistic Information Retrieval