

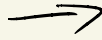
# KNN Classifier

also counts  
frequency

Constructing the Matrix

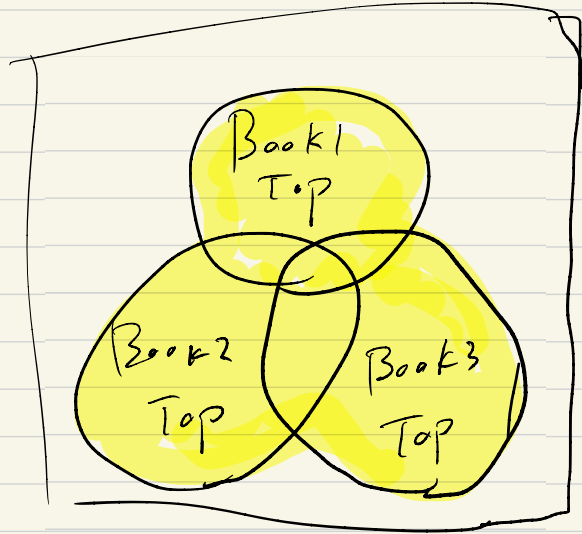


title  
author  
text



Top 100/200  
unique words

Book1 - top 100/200  
Book2 - top 100/200  
Book3 - top 100/200



take the set that  
contains all words

author

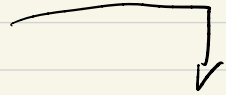
	"thou"	"quaint"	"pal"	"cga"	...	"Big Brother"	CLASS
Romeo and Juliet	199	263	68	436	...	0	1 = Shakespeare
To Kill a Mockingbird	18	1	0	2	...	0	2 = Harper Lee
1984	0	0	1	0	...	42	3 = George Orwell
...							

= Book123 MATRIX

Using KNN

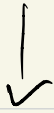


Top 100/200 words



	"thou"	"quaint"	...	CLASS
?	150	320	...	?
.				.

query-book



KNN ( query-book, book123\_matrix,  $k = 3$  )

↓ "magic"

Output : Shakespeare

Cosine Similarity

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

$\vec{a}$  = Book 123 MATRIX =  
n features

n books

	199	263	68	...	0
	18	1	0		0
	0	0	1		42
	⋮	⋮	⋮		⋮

$\vec{b}$  = query-book-vector<sup>T</sup> =

find  $\theta$  for each book in book<sub>123</sub> matrix

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} =$$

n features

$$\frac{\begin{bmatrix} 199 & 263 & 68 & \dots & 0 \end{bmatrix}}{\sqrt{199^2 + 263^2 + 68^2 + \dots + 0^2}} \cdot \frac{\begin{bmatrix} 150 \\ 320 \\ \vdots \\ \cdot \end{bmatrix}}{\sqrt{150^2 + 320^2 + \dots + \cdot^2}}$$

$$\theta = \cos^{-1} \left( \frac{\begin{bmatrix} \cancel{19} 9 & \cancel{26} 3 & \cancel{20} 8 & \dots & 0 \end{bmatrix} \begin{bmatrix} \cancel{15} 0 & 15 \\ \cancel{32} 0 & 32 \\ \vdots & \vdots \end{bmatrix}}{\sqrt{\cancel{19}^2 + \cancel{26}^2 + \cancel{20}^2 + \dots + 0^2} \sqrt{\cancel{15}^2 + \cancel{32}^2 + \dots^2}} \right)$$

just making the numbers easy to calculate with.

$$\theta = \cos^{-1} \left( \frac{20 \cdot 15 + 26 \cdot 32}{\sqrt{20^2 + 26^2} \sqrt{15^2 + 32^2}} \right)$$

$$\theta = \cos^{-1} \left( \frac{1132}{\sqrt{1076} \sqrt{1249}} \right)$$

$$\theta = \cos^{-1} \left( \frac{1132}{32.8 \cdot 35.3} \right)$$

$$\theta = \cos^{-1} (0.978)$$

GOOD SCORE!  
close to 1 which  
is what I wanted