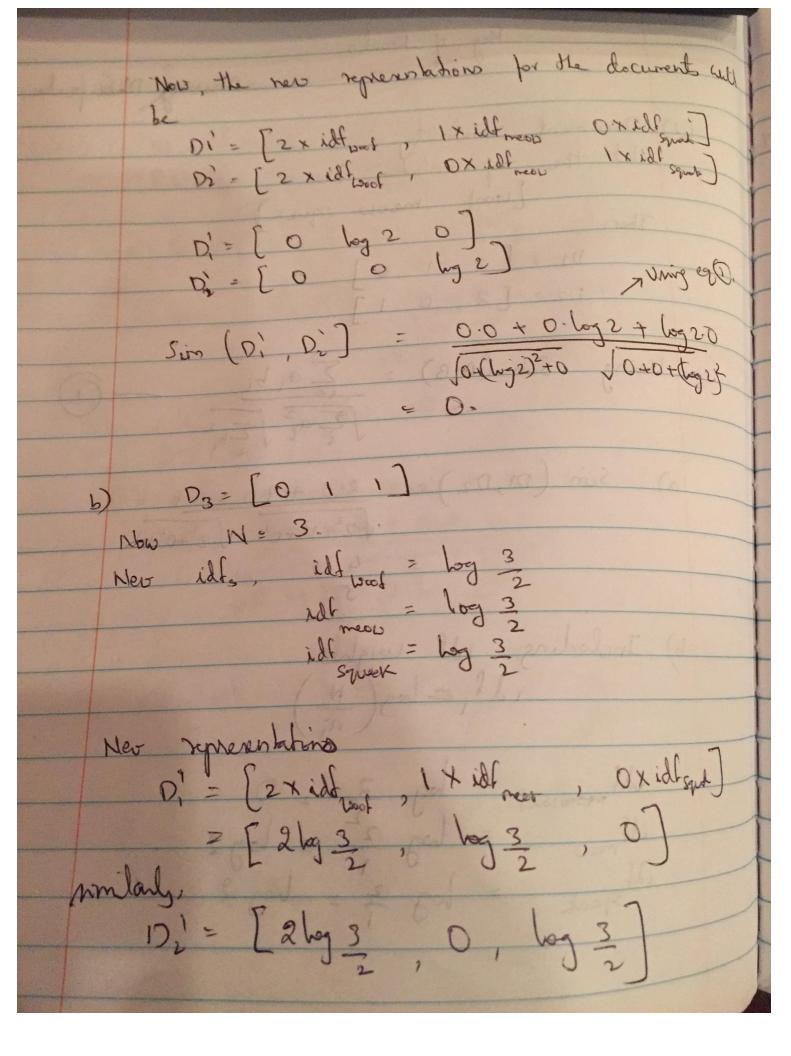
By of words -Anurag Dhaipule (1) Document Similarity
let the representation be
[woof means rquex] D1 = [2 1 0] D2 = [2 0 1 Uring Sim (A,B) = \(\sum_{io} a_i b_i \) 12 92 /2b2 a) Sim (D1, D2) = 2:2 + 1.0 + 0.1 $\sqrt{2^2+1^2+0^2}$, $\sqrt{2^2+0^2+1}$ b) Including edf beights $idf_i = log(\frac{N}{N})$ idf men = log 2 = 0

idf men = log 2 = log 2

idf men = log 2 = log 2

idf speak = log 2 = log 2.



Uring O, Sin (D', D') = 2 by 3/2 × 2 by 3/2 + 0+0 (2 kg 2) 2 (kg %) 20 (2 kg 3/2) 10+(kg 3/2) = 4(log 3/2) = 4 5(log 3/2)² = 5 (2) Naive Bayes and Smoothing.
(a) P(t) = Count of docs labelled to Tokel docs $P(+) = \frac{5}{10} = \frac{1}{2}$ $P(-) = \frac{1}{10} = \frac{1}{2}$ P(W.)): Can of does labelled & Containing Wi Coul of docs tabelled t. P(great |+) = 5 = 1 P(great 1-) = 0 = 0 P(food)+)=105) P(food)-) = 105 P(Sund 1+) = 0 = 0 P(servel) = 1 = 1 P(terible |-) = = 1 P(terrible |+) = = = 0

P(+1 great food server) = P(great 1+) x P(food 1+)x P(Soud)+) x (1-P(krible)+) xP(+) = 1×1×0×(1-0)×== P(-1" great food Served") = P(great 1-) x P(tood 1-) x P(Sured)-) x (1-P(kridul-) x P(-) = 0 * 1 * 1/ * (1-1) * 1/3 (b) with laplace smoothing,

we add

P(t) = Count of doc labelled t + 1

total doc + 2 $P(+) = \frac{6}{12} = \frac{1}{2}$ $P(-) = \frac{1}{2}$ P(great)+)= 5+1 - 6, P(great)-)= 1 5+2 7 P(food)+)= 5+1 , 6 ; P(food)-) = 6 P(Sewed) +) = 0+1 - 1; P(Somed) -) = 2 P(terribl)+): 0+1 = 1 P(sand)-)= 6

P(+) "grad food serve") = 6 x 6 x 1 P(+1" great food Served) is greater.