IS622 HW1

Brian Chu | Sept 1, 2015

Exercise 1.2.1: Using the information from Section 1.2.3, what would be the number of suspected pairs if the following changes were made to the data (and all other numbers remained as they were in that section)?

(a) The number of days of observation was raised to 2000.
The pairs of days would change to $\binom{2000}{2}$ or approximately $\frac{2000^2}{2} = 2 * 10^6$.
The resulting total is $5 * 10^{17} * 2 * 10^6 * 10^{-18} =$ 1,000,000 pairs.

(b) The number of people observed was raised to 2 billion (and there were therefore 200,000 hotels).
The chance they will visit the same hotel is now $\frac{.0001}{2*10^5} = 5 * 10^{-10}$.
The chance they will visit the same hotel on two different days is $(5 * 10^{-10})^2 = 5 * 10^{-20}$.
The number of pairs of people is also now $\frac{(2*10^9)^2}{2} = 2 * 10^{18}$.
The resulting total is $2 * 10^{18} * 5 * 10^5 * 5 * 10^{-20} =$ 50,000 pairs.

(c) We only reported a pair as suspect if they were at the same hotel at the same time on three different days.
The chance two people will meet on 3 days is $\binom{1000}{3}$ or approximately $1.7 * 10^9$.
The resulting total is $5 * 10^{17} * 1.7 * 10^9 * 10^{-18} = 8.5 * 10^8$ pairs.

Exercise 1.3.2: Suppose there is a repository of ten million documents, and word w appears in 320 of them. In a particular document d, the maximum number of occurrences of a word is 15. Approximately what is the TF.IDF score for w if that word appears (a) once (b) five times?

N = 10^7
w = 320
$IDF_w = log_2(\frac{10^7}{320}) = 14.9315686$

(a) TF = 1/15
TF.IDF = 1/15 * 14.9315686 = 0.9954379

(b) TF = 1/3
TF.IDF = 1/3 * 14.9315686 = 4.97719

Exercise 1.3.5: Use the Taylor expansion of e^x to compute, to three decimal places:

(a) $e^{1/10} = 1 + 0.1 + \frac{0.1^2}{2} + \frac{0.1^3}{6} + \frac{0.1^4}{24} + ...$
$= 1.105$

(b) $e^{-1/10} = 1 + -0.1 + \frac{-0.1^2}{2} + \frac{-0.1^3}{6} + \frac{-0.1^4}{24} + ...$
$= 0.905$ (inverse of above)

(c) $e^2 = 1 + 2 + \frac{2^2}{2} + \frac{2^3}{6} + \frac{2^4}{24} + \frac{2^5}{120} + \frac{2^6}{720} + \frac{2^7}{5040} + \frac{2^8}{40320} + ...$
$= 7.387$