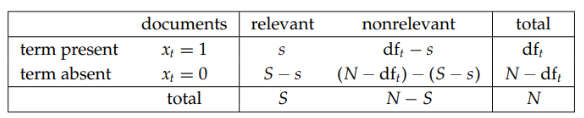
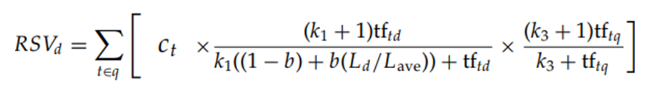
Q1: Which of these feature weighting techniques is best to give lower weights to words that occurs very frequency in all documents? (Give one line reason) [2 points]

1. Binary occurrence
2. Term Frequency
3. Term count
4. TF-IDF

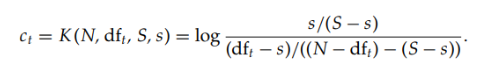
Q2: What is the underlying assumption in Vector space model? (Answer in 1 to 2 lines) [2 points]

Q3: Consider the following contingency table and retrieval status value (RSV) formula. \* [2 points]





Where ct is



Change the above formula if we assume that all the document containing term t are relevant.