Source task and conventions

Interpret every entry of $C^\top C$.

Conventional solution and proof

Entry (j, l) is $\sum_i C_{ij} C_{il}$, the number of vocabulary terms present in both documents j and l. On the diagonal it counts distinct represented terms in that document. For the two-document example in Equation (18.12), $C^\top C = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$. If repeated occurrences are encoded as counts, the same product becomes an unnormalized term-frequency inner product and loses the simple set-cardinality interpretation.