Source task and conventions

Compute the specified $CC^\top$ and interpret its large entries.

Conventional solution and proof

With $C = \left(\begin{smallmatrix} 0 & 2 & 1 \\ 0 & 3 & 0 \\ 2 & 1 & 0 \end{smallmatrix}\right)$,

$$CC^\top = \begin{pmatrix} 5 & 6 & 2 \\ 6 & 9 & 3 \\ 2 & 3 & 5 \end{pmatrix}.$$

For instance, terms 1 and 2 contribute $2 \cdot 3 = 6$ from document 2, whereas terms 1 and 3 contribute $2 \cdot 1 = 2$. Diagonal entries are sums of squared counts. The product weights co-occurrence by magnitude; a large diagonal can therefore reflect large raw counts and is not a normalized similarity score.

Structural or analytic derivation

Writing the product as a sum of document outer products $CC^\top = \sum_j c_j c_j^\top$ localizes each document's contribution and proves PSD simultaneously.

Comparative cost and preferred method

The three-column calculation is direct. In a sparse collection, outer-product accumulation can exploit short documents, but the total work depends on the sum of squared document nonzero counts. The Gram interpretation supplies correctness, not an unconditional sparse speedup.