

TASK 1

Compute the document-document similarity matrix, $A^t A$. (1 point)

Compute the term-term similarity matrix, AA^t . (1 point)

In words, what do these matrices mean? Qualitatively evaluate the numbers. (1 point)

TASK 2

Perform singular value decomposition (SVD) on the term-document matrix¹. (3 points)

Using your decomposition, compute the document-document similarity matrix, $(SD^t)^t(SD^t)$, and the term-term similarity matrix, $(TS)(TS)^t$. (1 point)

Compare the quality of these matrices to that of the matrices obtained in Task 1. (1 point)

TASK 3

Remove all but the first two singular values from your decomposition. Hence, \hat{T} will be a 12×2 matrix, \hat{S} will be a 2×2 matrix, and \hat{D} will be a 9×2 matrix. Using these, compute the document-document similarity matrix, $(\hat{S}\hat{D}^t)^t(\hat{S}\hat{D}^t)$, and the term-term similarity matrix, $(\hat{T}\hat{S})(\hat{T}\hat{S})^t$. (1 point)

Compare the quality of these matrices to that of the matrices obtained in Task 1 and Task 2. (1 point)

¹For python, you may use the SVD package within NumPy described here:
<http://docs.scipy.org/doc/numpy/reference/generated/numpy.linalg.svd.html>.