

Problem 1

- 1) Input is nothing must be run from the file containing document/queries
- 2) This prints in no particular order from the python file
- 3) For the most similar d/q pair it was unsurprising that two different method reported the same. However it appears the cosine measure did a better job of weeding out documents with the same word repeated many times.

Problem 2

$$(\lambda I_d + X^T X)^{-1} X^T$$

$$= \lambda I_d + (U \Sigma V^T)^T (U \Sigma V^T)$$

$$= \lambda I_d + (V^T \Sigma U^T) (U \Sigma V^T)$$

$$= (\lambda I_d + V \Sigma^2 V^T)^{-1} (U \Sigma V^T)$$

$$V(\lambda I_d) V^T + V \Sigma^2 V^T$$

$$V(\lambda I_d + \Sigma^2) V^T$$

$$(V S V^T)^{-1}$$

$$(V S^{-1} V^T) V \Sigma U^T$$

$$= V S^{-1} I_d \Sigma U^T.$$

$$= V (\lambda I_d + \Sigma^2) \Sigma U^T$$

$$\text{Let } S = \lambda I_d + \Sigma^2$$