1.

(a) R R N N N N N R N R N N N R N N N N R

recall | 0.125 0.25 0.375 0.5 0.625 0.75

precision | 1 1 0.33 0.364 0.33 0.3

$\Rightarrow P_{interp}(0.33) = \max_{r' \geq 0.33} p(r') = 0.364$

(b) $(1 + 1 + 0.33 + 0.364 + 0.33 + 0.3) / 8 = 0.4163$

(c) largest $\Rightarrow$ the rest 2 relevant docs are in rank 21th and 22nd.

$\Rightarrow$ MAP = $(1 + 1 + 0.33 + 0.364 + 0.33 + 0.3 + \frac{7}{21} + \frac{8}{22}) / 8 = 0.5034$

smallest $\Rightarrow$ They are in rank 9999th and 10000th.

$\Rightarrow$ MAP = $(1 + 1 + 0.33 + 0.364 + 0.33 + 0.3 + \frac{7}{9999} + \frac{8}{10000}) / 8 = 0.4165$

2.

(a) Use "accuracy" can measure the difference

$$\frac{\text{true positive} + \text{true negative}}{\text{collection size}}$$

(b) assume there are 20 relevant documents, and in 20 document the system retrieved, there are 10 relevant ones.

$\Rightarrow acc_{before} = \frac{tp + tn}{10000} = \frac{10 + ((10000 - 20) - (20 - 10))}{10000} = 0.998$

$acc_{after} = \frac{10 + ((10000 + 5000 - 20) - (20 - 10))}{10000 + 5000} = 0.99867$

$\Rightarrow$ Accuracy reflects the system performs better in the new setting.

C_{ij}, $i, j = 0 \ldots n-1$

default : $\{ (C_{00} + \cdots + C_{0n-1}) \ldots ((C_{n0-1} + \cdots + C_{n-1n-1}) = 0$