1) a) Raw term frequency doesn't account for how big the document is in terms of the number of words. In addition, stop words will occur more frequently but they should be equally likely among most documents in the corpus. Therefore, a better measure would be the term-frequency - inverse document frequency which reflects how important a word is with respect to other documents in the corpus. Usually, the cosine simmilarity is used rather than the dot product because the magnitude doesn't matter. so cosine simmilarity normalizes the document length of the two vectors that we are comparing.

b) Suppose the corpus that we're interested in has X instances of the term. Originally if $IDF = \log \frac{N}{|\{d \in D : t \in d\}|} = \log \frac{N}{B}$

then now $IDF = \log \frac{N+1}{B+x}$, so the IDF would decrease since the term is now more common in the corpus since we made a copy of d.

c) $N = 16$

	Retrieved	Not Retrieved
Relevant	5	6
Not Relevant	5	0 ↖ all 16 docs are relevant

i) $P = \frac{5}{10} = \boxed{0.5}$

ii) $R = \frac{5}{11} = \boxed{0.4545}$

iii) $F_1 = \frac{2 \cdot \frac{5}{10} \cdot \frac{5}{11}}{\frac{5}{10} + \frac{5}{11}} = \boxed{0.476}$

Rank as follows

iv) + + − + + − − + − −

1 1 0 $\frac{3}{4}$ $\frac{4}{5}$ 0 0 $\frac{5}{8}$ 0 0 $\Rightarrow$ sum $= 4.175 \xrightarrow{\div 16}$ avp $= 0.26$

d) i) $\boxed{CG = 5}$

$IDCG = 1 + 1 + \frac{1}{\log_2 3} + \frac{1}{2} + \frac{1}{\log_2 5} = 3.562$

ii) $DCG = 1 + 1 + \frac{1}{\log_2 4} + \frac{1}{\log_2 5} + \frac{1}{\log_2 8} = 3.264$

$\boxed{NDCG = \frac{3.264}{3.562} = 0.916}$

2)

a)

V	$p(A=1 \mid V)$	$p(K=1 \mid V)$	$P(L=1 \mid V)$	$P(V)$
0	2/6	3/6	2/6	1/2
1	4/6	5/6	2/6	1/2

b) $p(V=1 \mid A=0, K=1, L=0) = P(V=1 \mid A=0) \cdot P(V=1 \mid K=1) \cdot P(V=1 \mid L=0)$

$$= \frac{P(A=0 \mid V=1)\,P(V)}{P(A)} \cdot \frac{P(K=1 \mid V=1)\,P(V)}{P(K=1)} \cdot \frac{P(L=0 \mid V=1)\,P(V=1)}{P(L=0)}$$

$1-\frac{2}{6}$

$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$

$$= \frac{2/6 \cdot 1/2}{1/2} \cdot \frac{5/6 \cdot 1/2}{8/12} \cdot \frac{4/6 \cdot 1/2}{8/12}$$

$$= \frac{5}{48} = 0.104$$

$P(V=0 \mid A=0, K=1, L=0) = P(V=0 \mid A=0) \cdot P(V=0 \mid K=1) \cdot P(V=0 \mid L=0)$

$P(V=0 \mid A=0, K=1, L=0) + P(V=1 \mid A=0, K=1, L=0)$ should be 1

$$= \frac{P(A=0 \mid V=0)\,P(V=0)}{P(A=0)} \cdot \frac{P(K=1 \mid V=0)\,P(V=0)}{P(K=1)} \cdot \frac{P(L=0 \mid V=0)\,P(V=0)}{P(L=0)}$$

$\because P(V=0 \mid A=0, K=1, L=0) > P(V=1 \mid A=0, K=1, L=0)$

Therefore the message most likely does not contain a virus

$1-\frac{3}{6}$

$$= \frac{4/6 \cdot 1/2}{1/2} \cdot \frac{3/6 \cdot 1/2}{8/12} \cdot \frac{4/6 \cdot 1/2}{8/12}$$

$$= \frac{1}{8} = 0.125$$

c) If computed in the direct way as in a) then

$p(V=1 \mid A=0, K=1, L=0) = 0.5$, $P(V=0 \mid A=0, K=1, L=0) = 0.5$

this is because the pair $(A, K, L) = (0, 1, 0)$ is treated as one event but rather in the b) calculation they are treated as separate events

d) For the prior $p(V=0)$ and $p(V=1)$ must sum to 1 so we know that $p(V=1) = 1/2$. Other than that the posteriors should be arbitrary

2) e) Changing the A value in sample #0 would mean that $p(V=1 \mid A=0)$ would change so the $P(V=1 \mid A=0, K=1, L=0)$ result would change.

f) $p(A, L, K \mid V) = p(A \mid V)\, p(L \mid V)\, p(K \mid V)$

$$= \frac{p(V \mid A)\, p(A)}{p(V)} \; \frac{p(V \mid L)\, p(L)}{p(V)} \; \frac{p(V \mid K)\, p(K)}{p(V)}$$

We need to specify only two postoriors $\because p(V \mid A) + p(V \mid L) + p(V \mid K) = 1$

then we need to specify 4 priors $p(A), p(L), p(K), p(V)$

So we need 6 probability values as a minimum.

g) The events A, L, K may not be completely independent. For example, maybe messages that are shorter than 10 words are more likely to contain an attachment because the content is in the attachment so that the messages simply refer to the documents. So if event A & L are dependent, then the independence assumption $p(A, L, K \mid V) = p(A \mid V)\, p(L \mid V)\, p(K \mid V)$ is not completely valid

3) a) $l = p(X = \{x_1 \ldots x_n\}) = \frac{u^{x_1} e^{-u}}{x_1!} \cdot \frac{u^{x_2} e^{-u}}{x_2!} \cdot \ldots \frac{u^{x_n} e^{-u}}{x_n!}$

$$l = \frac{u^{\sum_i x_i} e^{-nu}}{\prod_i x_i!} \Rightarrow \mathcal{L} = \log l = \sum x_i \log u - nu \cancel{\log e} - \log \prod x_i!$$

$$\mathcal{L} = \log u \sum x_i - nu - \log \prod x_i!$$

$$\frac{\partial \mathcal{L}}{\partial u} = \frac{\sum x_i}{u} - n = 0 \Rightarrow \underbrace{\hat{u} = \frac{\sum x_i}{n}}_{}, $$

Best parameter is sample average

b) $l = p(U = \{u_1, \ldots u_n\}) = \lambda e^{-\lambda u_1} \cdot \lambda e^{\lambda u_2} \cdot \ldots \lambda e^{\lambda u_n}$

$$l = \lambda^n e^{-\lambda \sum_i u_i}$$

$$\mathcal{L} = \log l = n \log \lambda - \lambda \sum_i u_i \cancel{\log e}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{n}{\lambda} - \sum_i u_i = 0 \Rightarrow \left| \hat{\lambda} = \frac{n}{\sum_i u_i} \right|$$