1. Given the following data which describes three features (words) in a document (*Orange*, *School*, *Apple*) and the document outcome (whether its relevant R=1 or non-relevant NR=0)

	Orange	School	Apple	R or NR
Doc1	1	1	1	0
Doc2	1	1	0	0
Doc3	0	0	0	0
Doc4	0	1	0	1
Doc5	1	0	1	1
Doc6	0	1	1	1

Each row indicates the terms that were found in the document (1= appear, 0= not appear) and whether the document was relevant or not)

Use the Naive Bayes classifier to this data set and compute $P(R = 1|DOC_{new})$ which is represented by (1,0,0), i.e. assume that you have a document that contains the word *Orange* but does not contain *School* or *Apple*. I will use the notation "not Orange" to mean that we are looking at feature Orange to be equal to 0 in a document. So, instead of using the the terminology Orange=1 and Orange=0, I will use either "Orange" or "not Orange".

So, use Naive Bayes to classify if the document is relevant or non-relevant.

Naive Bayes says that the $P(R = 1|Doc_{new}) = P(R = 1|Orange) * P(R = 1|notSchool) * P(R = 1|notApple)$.

So, we will use Bayes rule to compute $P(R = 1|Orange)$, $P(R = 1|notSchool)$, and $P(R = 1|notApple)$. Bayes rule is $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Mush show $P(R = 1|Doc_{new}) = P(R = 1|Orange) * P(R = 1|notSchool) * P(R = 1|notApple)$, since Doc_{new} has word Orange but not the words School or Apple.

$P(R = 1|Orange) = \frac{P(Orange|R=1)P(R)}{P(Orange)} = \frac{\frac{1}{3}*\frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}$

$P(R = 1|notSchool) = \frac{P(notSchool|R=1)P(R)}{P(notSchool)} = \frac{\frac{1}{3}*\frac{1}{2}}{\frac{2}{6}} = \frac{1}{2}$

$P(R = 1|notApple) = \frac{P(notApple|R=1)P(R)}{P(notApple)} = \frac{\frac{1}{3}*\frac{1}{2}}{\frac{1}{3}} = \frac{1}{3}$

$P(R = 1|Doc_{new}) = P(R = 1|Orange) * P(R = 1|notSchool) * P(R = 1|notApple) = \frac{1}{3} * \frac{1}{2} * \frac{1}{3} = \frac{1}{18}$

So to decide whether a document is relevant or non-relevant. Now we would need to compute the $P(R = 0|Doc_{new})$ and do this all over again for R=0. Then we would compare $P(R = 0|Doc_{new})$ to $P(R = 1|Doc_{new})$ to see which one has higher probability.