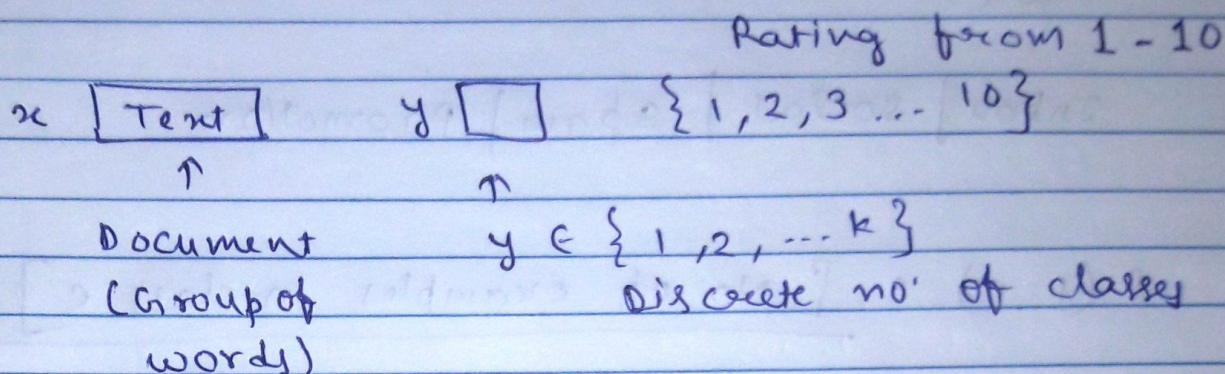


## ⑩ Naive Bayes for Text Classification

Movie Rating



$m$  documents  
(training reviews)

Prediction

$$P(y = c | x) \propto P(x | y = c) P(y = c)$$

$y$  = that class given  $x$  review

Naive Bayes :

Two assumption here :

① Bag of words

(order of words doesn't matter)

② Conditional Independence

$$P(x_i | y)$$

$$P(x_i | y = \text{true})$$

$$P(x_j | y = \text{true})$$

Happy

true

Liked

true

Independent of each other



$$x = \langle x_1, x_2, \dots, x_i, \dots, x_{|V|} \rangle$$

$$P(y=c | x_i) = \prod P(x_i | y=c) P(y=c)$$

Mail

Inbox | social | spam | promotion

$$P(y=c) = \frac{[\text{No. of examples in class } c]}{[\text{Total examples}]}$$

$x_i \rightarrow \text{"Happy"}$

$y \rightarrow \text{"+ve"}$

$$P(x_i | y=c) = \frac{[\text{count}(x_i, y_i=c)]}{\sum \text{count}(w, y_i=c)}$$

$w = \text{words in a particular class}$

denominator = Sum of counts of each word in vocab in class  $c$

numerator = Count of word  $x_i$  that occurred in class  $C$ .