Chapter No. 13 Text Classification and Naïve Bayes
Chapter No. 14 Vector Space Classification

<Food for Thoughts>

1. Define the general problem of text classification? Provide a mathematical model for it as well.
2. Differentiate between Multinomial Naïve Bayes-classifier and Bernoulli Naïve Bayes –classifier?
3. Define supervised learning? What are the essential requirements for doing supervised learning? Explain.

4. How Rocchio's Algorithm can be used for text classification? Explain it modification?
5. What are some of the weaknesses of Rocchio classification?
6. What are some of the benefits of using kNN for text classification?
7. Consider the following examples for the task of text classification

	docID	words in document	in $c = China$?
training set	1	Taipei Taiwan	yes
	2	Macao Taiwan Shanghai	yes
	3	Japan Sapporo	no
	4	Sapporo Osaka Taiwan	no
test set	5	Taiwan Taiwan Sapporo	?

a. Using the training data first calculate the class prior probabilities?
b. Using Multinomial Naïve Bayes to estimate the probabilities of each term (feature), that you will be using for doing part c?
c. Apply the Multinomial Naïve Bayes to classify the given test instance?