**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**BITS C464 – MACHINE LEARNING**

**I Semester 2014-2015**

**WORKSHEET #6**

**Naïve Bayesian Classifier**

**OBJECTIVE:-**

* **Bayesian Classifier**
* **Naïve Bayesian Classifier**
* **Classification using Naïve Bayesian Classifier**

**Bayesian Classifier:**

Consider a supervised learning problem in which we wish to approximate an

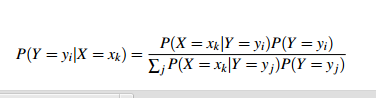
Unknown target function f : X → Y, or equivalently P(Y|X). To begin, we will

assume Y is a Boolean-valued random variable, and X is a vector containing n

boolean attributes. In other words, X = hX1,X2 ...,Xni, where Xi

is the Boolean random variable denoting the ith attribute of X.

Applying Bayes rule, we see that P(Y = yi | X) can be represented as



where ym denotes the mth possible value for Y, xk denotes the kth possible vector value for X, and where the summation in the denominator is over all legal values of the random variable Y.

One way to learn P(Y|X) is to use the training data to estimate P(X|Y) and

P(Y). We can use these estimates, together with Bayes rule above, to determine P(Y| X=Xk) for any instance Xk.

**Learning using Bayes classifier:**

Consider the following vector:

(likes shortbread, likes lager, eats porridge, watched England play football, nationality)T

A vector would describe that a person likes shortbread, does not like lager, eats porridge, has not watched England play football and is a national of Scottland. The final point is the class that we want to predict and takes two values: 1 for Scottish, 0 for English.

Here’s the data we’re given:

X = [ 0 0 1 1 0 ;  
1 0 1 0 0 ;  
1 1 0 1 0 ;  
1 1 0 0 0 ;  
0 1 0 1 0 ;  
0 0 1 0 0 ;  
1 0 1 1 1 ;  
1 1 0 1 1 ;  
1 1 1 0 1 ;  
1 1 1 0 1 ;  
1 1 1 1 1 ;  
1 0 1 0 1 ;  
1 0 0 0 1 ];

Notice that usually when we represent data, we write features in columns, instances in rows. If this is the case, we need to get the data in proper orientation: features in rows, instances in columns. Also, we need to separate the class from the feature set:

Y = X(:,5);

X = X(:,1:4)'; % X in proper format now.

In order to predict the class from a feature set, we need to find out the probability of Y given X (where



with n being the number of features. We denote the number of instances given to us as m. In our example, n = 4, m = 13. The probability of Y given X is:



Which is called the Bayes rule. Now, we make the NB assumption: All features in the feature set are independant of each other. Given this assumption, we need to find



To find , you just have to find  for all features and multiply them together. This is where the assumption comes in. You need the assumption of independence here for this.



This equation basically means count the number of instances for which both x\_i and Y are 1 and divide by the count of Y being 1. That’s the probability of x\_i appearing with Y. Fairly straight forward.



Same as above. Count the ratio of Y=1 with the total number of Ys. Notice that we need to calculate all these for both Y=0 and Y=1 because we need both in the first equation. For all of below, consider E as 0 and S as 1 since we consider being Scottish as being in class 1 (positive example).

**Exercise:**

1. **Implement a Bayes classifier by using the algorithm described above (use the dataset given in the example for training).**