**ASSIGNMENT 1**

**Naive Bayes Classifier**

Machine Intelligence and Expert System (IT-5213)

**Submitted By:**

|  |  |  |
| --- | --- | --- |
| Name | Roll | M.Sc. Session |
| 1. Tajim Md. Niamat Ullah Akhund | 1120 | 2017-2018 |
| 1. M. Mahfuzul Haq | 1124 |
| 1. Priyanka Dhar | 1074 |
| 1. Rafia Akther | 1086 |
| 1. Kazi Zannath Nowshin | 1077 |

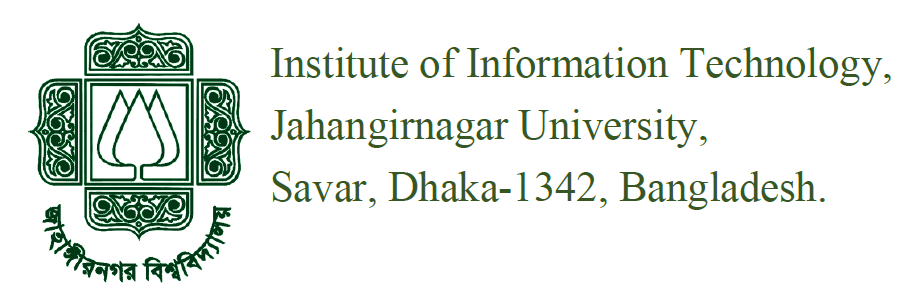
**Submitted To:**

**Dr. Shamim Al Mamun**

Associate Professor,

Institute of Information Technology,

Jahangirnagar University.



**Naive Bayes Classifier**

**Naive Bayes:**

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems [1].

Naive Bayes algorithm is based on Bayesian Theorem.

**Bayesian Theorem:**

Given training data X, posterior probability of a hypothesis H, P(H|X), follows the Bayes theorem

P(H|X) = {P(X|H) \* P(H)} / P(X) …………………………. (1.1)

**Algorithm:**

The Naive Bayes algorithm is based on Bayesian theorem as given by equation (1.1)

Steps in algorithm are as follows [2]:

1. Each data sample is represented by an n dimensional feature vector, X = (x1, x2….. xn), depicting n measurements made on the sample from n attributes, respectively A1, A2, An.
2. Suppose that there are m classes, C1, C2……Cm. Given an unknown data sample, X (i.e., having no class label), the classifier will predict that X belongs to the class having the highest posterior probability, conditioned if and only if:

P(Ci/X)>P(Cj/X) for all 1< = j< = m and j != i

Thus, we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis.

1. As P(X) is constant for all classes, only P(X|Ci)P(Ci) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e. P(C1) = P(C2) = …...= P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class prior probabilities may be estimated by P(Ci) = si/s, where Si is the number of training samples of class Ci, and s is the total number of training samples. on X. That is, the naive probability assigns an unknown sample X to the class Ci [2]

**Example of Naive Bayes Classifier:**

We can predict the class of an animal from some attributes of it. Lets our train data set is as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **Give Birth** | **Can Fly** | **Live in Water** | **Have Legs** | **Class** |
| human | yes | no | no | yes | mammals |
| python | no | no | no | no | non-mammals |
| salmon | no | no | yes | no | non-mammals |
| whale | yes | no | yes | no | mammals |
| frog | no | no | sometimes | yes | non-mammals |
| komodo | no | no | no | yes | non-mammals |
| bat | yes | yes | no | yes | mammals |
| pigeon | no | yes | no | yes | non-mammals |
| cat | yes | no | no | yes | mammals |
| leopard shark | yes | no | yes | no | non-mammals |
| turtle | no | no | sometimes | yes | non-mammals |
| penguin | no | no | sometimes | yes | non-mammals |
| porcupine | yes | no | no | yes | mammals |
| eel | no | no | yes | no | non-mammals |
| salamander | no | no | sometimes | yes | non-mammals |
| qila monster | no | no | no | yes | non-mammals |
| platypus | no | no | no | yes | mammals |
| owl | no | yes | no | yes | non-mammals |
| dolphin | yes | no | yes | no | mammals |
| eagle | no | yes | no | yes | non-mammals |

Lets,

**A**: attributes

**M**: mammals

**N**: non-mammals

Then the equations will be:

Lets,

A given data is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Give Birth** | **Can Fly** | **Live in Water** | **Have Legs** | **Class** |
| yes | no | yes | no | ? |

We have to predict the class of that animal from the given data.

Then,

**References:**

[1] Mai Shouman, Tim Turner, Rob Stocker, “Using data mining techniques in heart disease diagnosis and treatment”, JapanEgypt Conference on Electronics, Communications and Computers 978-1-4673-0483-2 c\_2012 IEEE.

[2] N. Aaditya Sunder, P. PushpaLatha, “Performance analysis of classification data mining techniques over heart disease database” Inernational Journal Of Engineering Science and Advance Technology”-vol-2 issue-3,470-478,May-June 2012.