Faculty of Electronics & Information Technology
Warsaw University of Technology

Intelligent Information Systems

Prof. M. Muraszkiewicz, *mrm@ii.pw.edu.pl*

Lecture Notes: Module 1A (draft)

Naive Bayes Classifier

Module 1A (draft)

Prologue

Thomas Bayes and his theorem

Thomas Bayes
1702-1761

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bon mots

All models are wrong, but some are useful.

–George E. P. Box (1919-2013)
A British statistician, who worked in the areas of quality control, time-series analysis, design of experiments, and Bayesian inference

An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.

–John Tukey(1915-2000)
An American mathematician best known for development of the FFT algorithm

Table of Contents

Data, Information, Knowledge

Three Levels

Data

1	woman	<=30	medium	higher ed.	Yes
2	man	31…40	high	primary	Yes
3	woman	>40	low	secondary	No
…	…	…	…	…	…

This is data–some strings; we do not know their meaning

Information (Information System)

Id	Sex	Age	Income	Education	Credit rating
1	woman	<=30	medium	higher	Yes
2	man	31…40	high	primary	Yes
3	woman	>40	low	secondary	No
…	…	…	…	…	…

This is information–the header row determines interpretation.
A row represents an assertion, e.g. row #3: This is a woman whose age is above 40 years old with a low income, secondary education, and no creditworthiness.

Knowledge (Intelligent Information System)

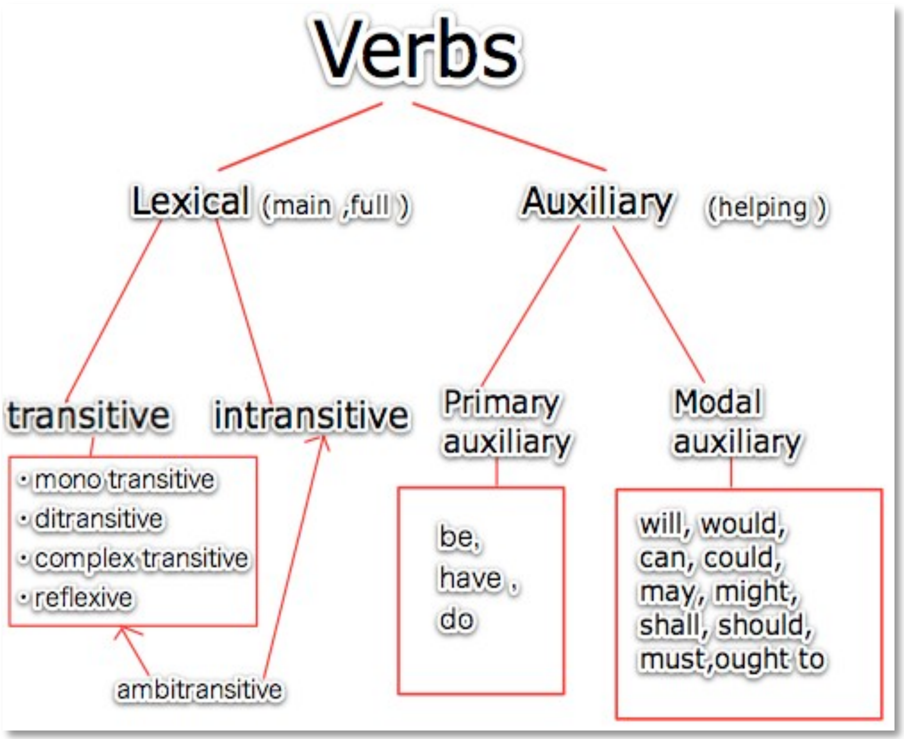
Id	Sex	Age	Income	Education	Credit rating
1	woman	<=30	medium	higher	Yes
2	man	31…40	high	primary	Yes
3	woman	>40	low	secondary	No
…	…	…	…	…	…

This is knowledge–A row represents a conditional statement IF-THEN, e.g. row#3:
IF this is a women above 40 years old and low income and secondary level of education, **THEN** she is not creditworthy.

Classification

Why to Classify?

People have always been classifying things (material and immaterial). Classifying is about assigning things to groups, classes, categories. We classify in order to better organize our life, to monitor changes, to compare the things, to examine them, to facilitate decision making, to …



Classification is a general process related to categorization, the process in which ideas and objects are recognized, differentiated, and understood. A classification system is an approach to accomplishing classification.

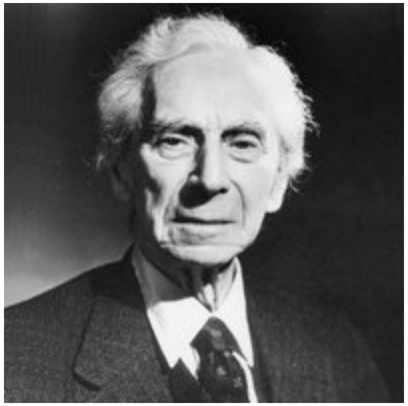
Classification and Knowledge

One who knows
can classify,

and

One who
classifies, knows

Z. Pawlak



B. Russel

Naïve Bayes Classifier

Features

This is why the classifier is called naive

Bayesian classification is based on the Bayes theorem and belongs to a family of simple probabilistic classifiers with strong (naive) independence assumptions between the features (attributes). In spite of its simplicity it offers a pretty good quality of classification. It not only indicates the class to which a given object is classified but also provides the probability of the assignment.

Its main drawbacks are that (i) from the user's standpoint it works as a black box, meaning it does not "explain" how the process of classification has been done; (ii) It assumes every feature is independent, which isn't always the case.

Task of Classification

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Given is a training set having n+1 attributes A_1, A_2, …, $A_{n,}$ A_{n+1}, where A_{n+1} is a decision attribute.

Let X = (x_1,x_2,...,x_n) be the object to be classified, where x_1,x_2,...,x_n take their values from the domains of the attributes A_1, A_2, …, A_n, respectively.

Classification goes in two steps:

1. Calculate $P(C_i|X)$ for all classes C_i belonging to C = (C_1, C_2, …, C_m), i.e. $P(C_1|X)$, $P(C_2|X)$, …, $P(C_m|X)$
2. Assign object X to the class for which $P(C_i|X)$ takes the maximum value, for I = 1, 2, …, m.

Task of Classification

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Note that the value of P(X) is constant for all classes. Thus, the C_i, for which $P(C_i|X)$ gets maximum, is the C_i, for which $P(X|C_i) \times P(C_i)$ takes maximum value.

$P(C_i) = \frac{z_i}{z}$, where z_i is the number of objects belonging to C_i and z is the number of all objects in the training set.

While calculating $P(C_i|X)$ we assume that **the attributes are independent of each other**, which allows us to calculate the values of $P(X|C_i)$, for $X = (x_1,x_2,...,x_n)$ as follows:

$$P(X|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i)$$

where

$$P(x_k|C_i) = z_{ik}/z_i \text{ dla } k = 1, 2, \ldots, n$$

z_{ik} is the naumber of objects in class C_i, for which the value of A_k is x_k and z_i is the number of all objects belonging to class C_i in a training set.

Training set

	Age	Income	Student	Credibility	Purchase
1	<=30	high	No	good	No
2	<=30	high	No	excellent	No
3	31...40	high	No	good	Yes
4	>40	medium	No	good	Yes
5	>40	low	Yes	good	Yes
6	>40	low	Yes	excellent	No
7	31...40	low	Yes	excellent	Yes
8	<=30	medium	No	good	No
9	<=30	low	Yes	good	Yes
10	>40	medium	Yes	good	Yes
11	<=30	medium	Yes	excellent	Yes
12	31...40	medium	No	excellent	Yes
13	31...40	high	Yes	good	Yes
14	>40	medium	No	excellent	No

C_1 = Yes, C_2 = No

Object to classify:

X = {Age <=30, Income = medium, Student = Yes, Credibility = good}

Let's calculate

Calculate the products $P(X|C_i) \times P(C_i)$, for $i = 1,2$ and get the one, which takes the maximum value.

To this end, using the training set calculate the following:

$P(C_1) = 9/14 = 0.643$

$P(C_2) = 5/14 = 0.357$

C_1 = Yes, C_2 = No

$P(X|C_1) = P(Age <=30|C_1) * P(Income = medium|C_1) * P(Student = Yes|C_1) * P(Credibility = good|C_1)$

$P(X|C_2) = P(Age <=30|C_2) * P(Income = medium|C_2) * P(Student = Yes|C_2) * P(Credibility = good|C_2)$

Object to classify:

X = {Age <=30, Income = medium, Student = Yes, Credibility = good}

Let’s calculate

$P(X|C_i) \times P(C_i)$, for i = 1,2

$P(X|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i)$

C_1 = Yes, C_2 = No

	Age	Income	Student	Credibility	Purchase
1	<=30	high	No	good	No
2	<=30	high	No	excellent	No
3	31...40	high	No	good	Yes
4	>40	medium	No	good	Yes
5	>40	low	Yes	good	Yes
6	>40	low	Yes	excellent	No
7	31...40	low	Yes	excellent	Yes
8	<=30	medium	No	good	No
9	<=30	low	Yes	good	Yes
10	>40	medium	Yes	good	Yes
11	<=30	medium	Yes	excellent	Yes
12	31...40	medium	No	excellent	Yes
13	31...40	high	Yes	good	Yes
14	>40	medium	No	excellent	No

$P(Age <=30|C_1) = 2/9 = 0.222$
$P(Income = medium|C_1) = 4/9 = 0.444$
$P(Student = Yes|C_1) = 6/9 = 0.667$
$P(Credibility = good|C_1) = 6/9 = 0.667$

$P(Age <=30|C_2)$) $= 3/5 = 0.600$
$P(Income = medium|C_2) = 2/5 =0.400$
$P(Student = Yes|C_2) = 1/5 = 0.200$
$P(Credibility = good|C_2) = 2/5 = 0.400$

$P(X|C_1)=0.222*0.444*0.667*0667=0.044$
$P(X|C_1) * P(C_1) = 0.044*0.643 = 0.028$

$P(X|C_2)=0.600*0.400*0.200*0.400=0.019$
$P(X|C_2) * P(C_2) = 0.019 * 0.357 = 0.007$

Object X = {Age <=30, Income = medium, Student = Yes, Credibility = good} has been classified to C_1.

A Naïve Bayes Classifier
is and intelligent information system!

Question:

What is a language
and a reasoning mechanism
in a Naïve Bayes Classifier?

Question:

What is a language
and a reasoning mechanism
in a Naïve Bayes Classifier?

Answer:

language: table (relation)
reasoning: Bayes formula

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs) STOP
Predicted Negative (0)	False Negatives (FNs) STOP	True Negatives (TNs)

A good classifier minimises the number of FN and FP errors

Measures of Classification Quality

$$Error\ rate = \frac{FN + FP}{TP + FN + FP + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

Sensitivity tells how good is the classifier to predict the actual state (TP).

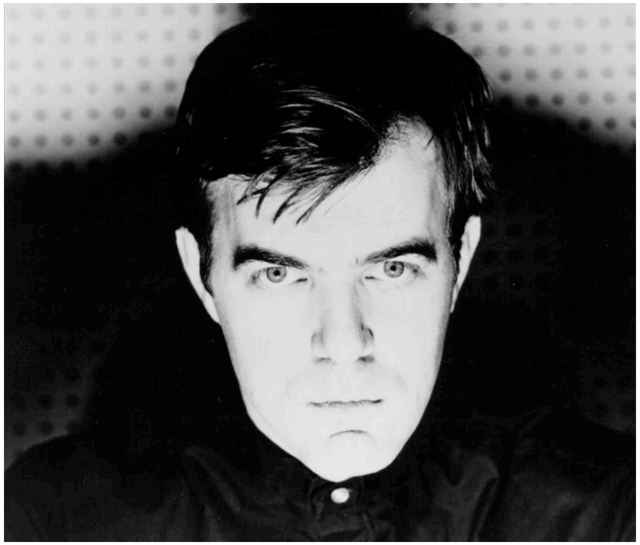
$$Specificity = \frac{TN}{TN + FP}$$

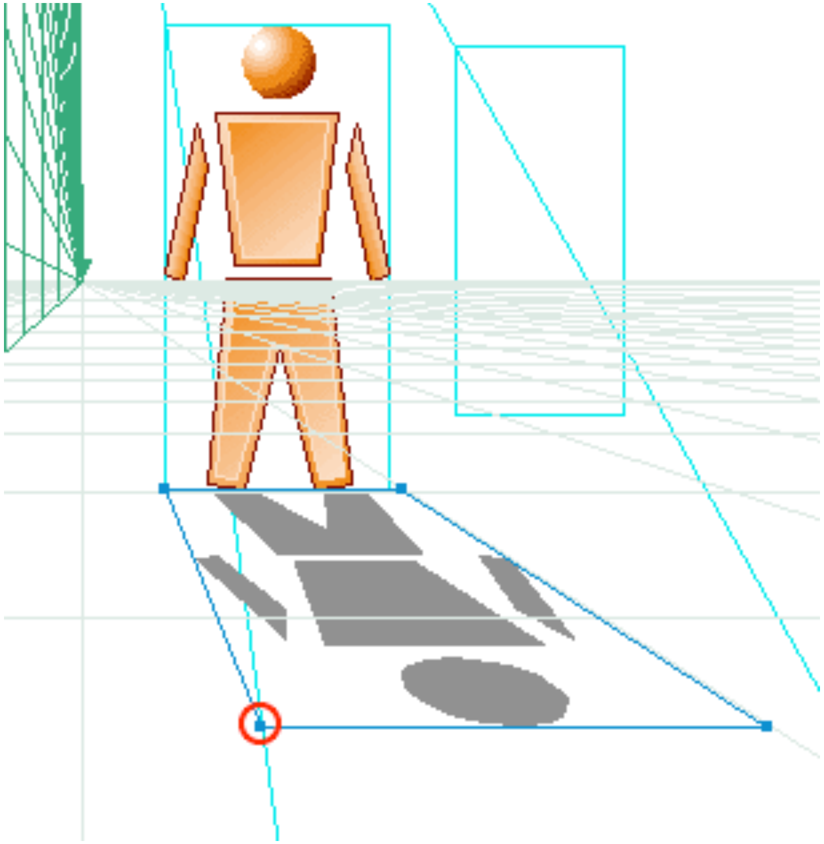
Specificity tells how effective is the classifier to predict the false state (TN).

Epilogue

To be beyond any existing classification has always pleased me.

–Boyd Rice





Thank you!