

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«Ижевский государственный технический университет  
имени М.Т. Калашникова»

Отчет по практической работе  
на тему «Наивный Байесовский классификатор»  
по дисциплине «Математическое моделирование»

Выполнил:  
студент гр. М01-191-1

Филиппов А.Н.

Проверил:

Тененев В.А.

Ижевск 2016

# Ход работы

Дан набор точек и определённые им классы.

x1	x2	x3	q
2	2	1	4
1	1	2	2
2	2	1	4
1	2	1	2
1	1	2	2
1	2	1	2
2	1	2	4
2	2	2	5
1	1	1	1
1	2	2	4
2	1	2	4
1	2	2	4
1	1	2	2
1	1	2	2
1	1	1	2
1	1	1	2
1	2	2	4
1	1	2	2
2	1	1	2
2	1	2	4
2	2	1	4
1	1	2	2

1	1	1	1
2	1	2	4
2	2	2	5
2	1	2	4
2	1	1	2
1	1	1	1
2	2	2	5
1	2	2	4

На первом этапе вычисляются количества значений  $x_j=1$  и  $H_q=q$ .

$H_{qj1}$						
$j=1$	$j=2$	$j=3$	$H_q$	$q$	$1$	
3	3	3	3	3	1	1
0	0	0	0	3	1	2
9	9	4	11	2	2	1
2	2	7	11	2	2	2
0	0	0	0	0	3	1
0	0	0	0	0	3	2
5	5	3	13	4	4	1
8	8	10	13	4	4	2
0	0	0	3	5	5	1
3	3	3	3	3	5	2

Используя полученные значения, вычисляются априорные вероятности появления каждого класса  $p(q) = Nq/N$  и параметра  $p(x_j | q) = N_{qj}/Nq$ .

p(x <sub>j</sub>   q)						
j=1	j=2	j=3	p(q)	q	1	
1	1	1	0.1	1	1	1
0	0	0	0.1	1	1	2
0.8181818181 82	0.8181818181 82	0.3636363636 36	0.3666666666 67		2	1
0.1818181818 18	0.1818181818 18	0.6363636363 64	0.3666666666 67		2	2
0	0	0	0		3	1
0	0	0	0		3	2
0.3846153846 46	0.3846153846 46	0.2307692307 08	0.4333333333 33		4	1
0.6153846153 54	0.6153846153 54	0.7692307692 92	0.4333333333 33		4	2
0	0	0	0.1		5	1
1	1	1	0.1		5	2

Используя полученные априорные вероятности можно построить наивный Байесовский классификатор. Ниже представлена классификация заданной точки (1, 1, 2) с его помощью.

Sample		
x1	x2	x3
1	1	2

Для классификации точки необходимо вычислить  $\text{argmax } p(q|x)$ . Это значит, что необходимо найти такое значение  $q$  для заданного  $x$ , что значение  $p(q|x)$  будет максимальным. Полученное значение  $q$  – это класс  $x$  с вероятностью

$p(q|x)$ . Для вычисления  $p(q|x)$  необходимо вычислить произведение априорных вероятностей появления как самого класса, так и каждого из параметров в отдельности  $p(q)*p(x1|q)*p(x2|q)*p(x3|q)$ , которые берутся из таблицы выше. Для удобства данные дублируются в результирующей таблице.

q	$p(q x)$	$p(x1   q)$	$p(x2   q)$	$p(x3   q)$	$p(q)$
1	0	1	1	0	0.1
2	0.1561983471	0.8181818182	0.8181818182	0.6363636364	0.3666666667
3	0	0	0	0	0
4	0.04930966469	0.3846153846	0.3846153846	0.7692307692	0.4333333333
5	0	0	0	1	0.1

В результате вычислений был определён класс точки (1, 1, 2)  $q=2$  с вероятностью  $p(2|(1, 1, 2))=0.156$ .

Работа выполнена при помощи инструмента Google Sheets – электронных таблиц компании Google. Все данные за исключением вводных вычисляются автоматически при помощи формул. Делается допущение, что входных образцов всегда 30, а параметров (координат) 3. Это существенно упрощает реализацию. В отчёт скопированы результаты вычислений для заданного примера.

## Вывод

Наивный Байесовский классификатор позволяет осуществить классификацию неизвестных данных по известной заранее выборке. В нём делается допущение, что параметры образца независимы друг от друга, такое допущение редко выполняется на практике, потому называется наивным. Основным преимуществом данного классификатора является его простота реализации. На практике он применяется как правило в качестве строительного блока для более сложных алгоритмов классификации.