P03 Probability Distributions and Bayesian Classification

Luis Sa-Couto[1] and Andreas Wichert[2]

INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa
{luis.sa.couto,andreas.wichert}@tecnico.ulisboa.pt

1 Probability and Distributions

1) Consider the following registry where an experiment is repeated six times and four events (A, B, C and D) are detected. Considering frequentist estimates for probabilities, compute:

	D	C	B	A
1	0	0	1	1
2	0	1	1	1
3	1	0	0	0
4	1	0	0	0
5	0	0	0	0
6	0	0	0	0

- $p(A)$
- $p(A, B)$
- $p(B \mid A)$
- $p(A, B, C)$
- $p(A \mid B, C)$
- $p(A, B, C, D)$
- $p(D \mid A, B, C)$

2) Consider the following set of height measures in centimeters of a group of people:

X					
180	160	200	171	159	150

What are the maximum likelihood parameters of a gaussian distribution for this set of points? Plot it approximately.

3) Consider the following set of two dimensional measures:

X_1	-2	-1	0	-2
X_2	2	3	1	1

What are the maximum likelihood parameters of a Gaussian distribution for this set of points? What is the shape of the Gaussian? Draw it approximately using a contour map.

4) Consider the following set of two dimensional measures:

$$\begin{array}{c|cccc} X_1 & 2 & 1 & 0 & 2 \\ X_2 & -2 & 3 & -1 & 1 \end{array}$$

What are the maximum likelihood parameters of a Gaussian distribution for this set of points? What is the shape of the Gaussian? Draw it approximately using a contour map.

2 Simple Bayesian Learning

1) Assuming that 1 means *True* and 0 means *False*, consider the following features and class:

- X_1: "Fast processing"
- X_2: "Decent Battery"
- X_3: "Good Camera"
- X_4: "Good Look and Feel"
- X_5: "Easiness of Use"
- *Class*: "iPhone"

You are given the following training set:

X_1	X_2	X_3	X_4	X_5	*Class*
1	1	0	1	0	1
1	1	1	0	0	0
0	1	1	1	0	0
0	0	0	1	1	0
1	0	1	1	1	1
0	0	1	0	0	1
0	0	0	0	1	1

And the query vector $\mathbf{x} = \begin{bmatrix} 1\,1\,1\,1\,1 \end{bmatrix}^T$.

a) Using Bayes' rule, without making any assumptions, compute the class for the query vector.

b) What is the problem of working with this data set without assumptions?

c) Compute the class for the same query vector under the Naive Bayes assumption.

2) From the following training set:

X_1	X_2	X_3	X_4	X_5	C
1	1	0	1	0	a
1	0	0	1	1	a
1	0	0	1	1	a
1	1	1	0	1	b
0	0	1	1	1	b
1	0	0	0	0	c

a) Compute the class for the pattern $\mathbf{x} = \begin{bmatrix} 1\ 0\ 1\ 0\ 1 \end{bmatrix}^T$ under the Naive Bayes assumption.

b) What is the posterior probability $p(b \mid \mathbf{x})$?

c) What do you do if we have missing features? More specifically, under the Naive Bayes assumption, to what class does $\mathbf{x}_{\mathbf{missing}} = \begin{bmatrix} 1\ ?\ 1\ ?\ 1 \end{bmatrix}^T$ belong to?

3) So far we have been dealing always with discrete feature domains. In this exercise, we will work with continuous values for features like Height and Weight. Assuming that 1means $True$ and 0means $False$, consider the following features and class:

- X_1: "Weight (Kg)"
- X_2: "Height (Cm)"
- $Class$: "NBA Player"

You are given the following training set:

X_1	X_2	$Class$
170	160	0
80	220	1
90	200	1
60	160	0
50	150	0
70	190	1

And the query vector $\mathbf{x} = \begin{bmatrix} 100\ 225 \end{bmatrix}^T$.

a) Compute the most probable class for the query vector assuming that the likelihoods are 2-dimensional Gaussians.

b) Compute the most probable class for the query vector, under the Naive Bayes assumption, using 1-dimensional Gaussians to model the likelihoods.

3 Thinking Questions

a) Assuming training examples with d boolean features, how many parameters do you have to estimate if you make no assumptions about how the data is distributed? What about if you make the Naive Bayes assumption?

b) Is Naive Bayes a linear classifier?