CS 440: Introduction to Artificial Intelligence

Lecture 14

Matthew Stone

October 26, 2015

Supervised categorization—Recap

- Infer category of real-world object from features
- Start from examples
- Learn decision boundary
- Apply learned rule to new cases

Understanding classification via probability—Recap

We want to decide the most likely category

- Compute $P(C = c_1 | O = o)$
- Compute $P(C = c_2 | O = o)$, etc.
- Pick whichever one is the largest

Allows us to describe the optimum decision boundary

Naive Bayes assumption—Recap

- Ignore certain kinds of interactions in world
- Lets you use same data to learn multiple relationships
- Mathematically:

$$P(F_i|C) = P(F_i|C, F_1 \ldots F_{i-1})$$

- As a result:

$$P(F_1 \ldots F_n|C) = P(F_1|C)P(F_2|C) \ldots P(F_n|C)$$

Back to our case study

We have this training data:

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

We will build a Naive Bayes model

What is step 1?

What is step 1?

Compute the priors

What is step 1?

Compute the priors

$$P(C = c_1)$$
$$P(C = g_2)$$

Compute the Priors

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$P(C = c_1) =$
$P(C = g_2) =$

Compute the Priors

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$$P(C = c_1) = 0.6$$
$$P(C = g_2) = 0.4$$

What is step 2?

What is step 2?

Compute the likelihoods

What is step 2?

Compute the likelihoods

$$P(T = w_1 | C = c_1)$$
$$P(T = f_2 | C = c_1)$$
$$P(T = w_1 | C = g_2)$$
$$P(T = f_2 | C = g_2)$$

and

$$P(S = h_1 | C = c_1)$$
$$P(S = t_2 | C = c_1)$$
$$P(S = h_1 | C = g_2)$$
$$P(S = t_2 | C = g_2)$$

Compute the Likelihoods

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$$P(T = w_1 | C = c_1) =$$
$$P(T = f_2 | C = c_1) =$$
$$P(T = w_1 | C = g_2) =$$
$$P(T = f_2 | C = g_2) =$$

Compute the Likelihoods

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$$P(T = w_1 | C = c_1) = .8$$
$$P(T = f_2 | C = c_1) = .2$$
$$P(T = w_1 | C = g_2) = .4$$
$$P(T = f_2 | C = g_2) = .6$$

Compute the Likelihoods

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$P(S = h_1 | C = c_1) =$
$P(S = t_2 | C = c_1) =$
$P(S = h_1 | C = g_2) =$
$P(S = t_2 | C = g_2) =$

Compute the Likelihoods

T	S	C	count
$T = w_1$	$S = h_1$	$C = c_1$	32
$T = f_2$	$S = h_1$	$C = c_1$	8
$T = w_1$	$S = t_2$	$C = c_1$	16
$T = f_2$	$S = t_2$	$C = c_1$	4
$T = w_1$	$S = h_1$	$C = g_2$	4
$T = f_2$	$S = h_1$	$C = g_2$	6
$T = w_1$	$S = t_2$	$C = g_2$	12
$T = f_2$	$S = t_2$	$C = g_2$	18

$$P(S = h_1 | C = c_1) = \frac{2}{3}$$
$$P(S = t_2 | C = c_1) = \frac{1}{3}$$
$$P(S = h_1 | C = g_2) = .25$$
$$P(S = t_2 | C = g_2) = .75$$

Classifying a new example

New item comes in: $T = f_2$, $S = h_1$.
What do you do?

$T = f_2, S = h_1$

Model:

$$
\begin{array}{lll}
 & P(T = w_1 | C = c_1) = .8 & P(S = h_1 | C = c_1) = \frac{2}{3} \\
P(C = c_1) = 0.6 & P(T = f_2 | C = c_1) = .2 & P(S = t_2 | C = c_1) = \frac{1}{3} \\
P(C = g_2) = 0.4 & P(T = w_1 | C = g_2) = .4 & P(S = h_1 | C = g_2) = .25 \\
 & P(T = f_2 | C = g_2) = .6 & P(S = t_2 | C = g_2) = .75
\end{array}
$$

What do we compute?

$T = f_2,\ S = h_1$

Model:

$$
\begin{array}{lll}
 & P(T = w_1 | C = c_1) = .8 & P(S = h_1 | C = c_1) = \frac{2}{3} \\
P(C = c_1) = 0.6 & P(T = f_2 | C = c_1) = .2 & P(S = t_2 | C = c_1) = \frac{1}{3} \\
P(C = g_2) = 0.4 & P(T = w_1 | C = g_2) = .4 & P(S = h_1 | C = g_2) = .25 \\
 & P(T = f_2 | C = g_2) = .6 & P(S = t_2 | C = g_2) = .75
\end{array}
$$

What do we compute?

- $P(C = c_1) \times P(T = f_2 | C = c_1) \times P(S = h_1 | C = c_1)$
- $P(C = g_2) \times P(T = f_2 | C = g_2) \times P(S = h_1 | C = g_2)$

$T = f_2$, $S = h_1$

Model:

$$P(C = c_1) = 0.6 \qquad P(C = g_2) = 0.4$$

$$P(T = w_1 | C = c_1) = .8 \qquad P(T = f_2 | C = c_1) = .2 \qquad P(T = w_1 | C = g_2) = .4 \qquad P(T = f_2 | C = g_2) = .6$$

$$P(S = h_1 | C = c_1) = \frac{2}{3} \qquad P(S = t_2 | C = c_1) = \frac{1}{3} \qquad P(S = h_1 | C = g_2) = .25 \qquad P(S = t_2 | C = g_2) = .75$$

What do we compute?

- $P(C = c_1) \times P(T = f_2 | C = c_1) \times P(S = h_1 | C = c_1) = .6 \times .2 \times \frac{2}{3} = .08$
- $P(C = g_2) \times P(T = f_2 | C = g_2) \times P(S = h_1 | C = g_2) = .4 \times .6 \times .25 = .06$

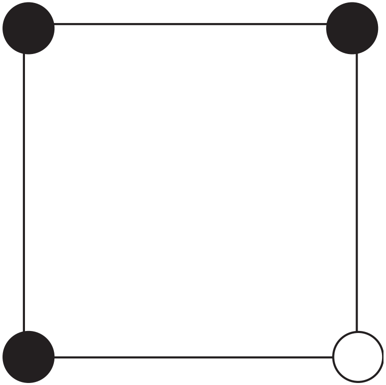
Evaluation Questions

- Is the Naive Bayes assumption justified in this example?
- What error do you expect for the learned classifier?

Depend on how estimated parameters relate to actual data.

Need for More General Algorithms

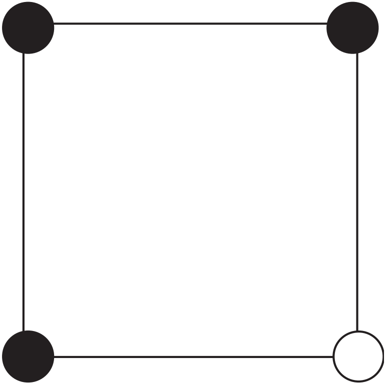
Intuition — simple case



We know what the decision boundary should be here

Need for More General Algorithms

Intuition — simple case

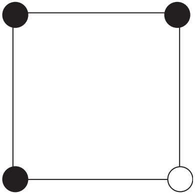


Naive Bayes might not find this
if independence assumptions are violated

Example

A	B	C	$P(A\&B\&C)$
0	0	0	.10
0	0	1	.15
0	1	0	.10
0	1	1	.15
1	0	0	.25
1	0	1	0
1	1	0	.10
1	1	1	.15

Supports classification rule:



Example

A	B	C	$P(A\&B\&C)$
0	0	0	.10
0	0	1	.15
0	1	0	.10
0	1	1	.15
1	0	0	.25
1	0	1	0
1	1	0	.10
1	1	1	.15

Naive Bayes parameters:

$P(C=0) = 11/20$

$P(C=1) = 9/20$

$P(A=0|C=0) = 4/11$

$P(B=0|C=0) = 7/11$

$P(A=0|C=1) = 6/9$

$P(B=0|C=1) = 3/9$

Example

Actual distribution:

A	B	C	$P(A\&B\&C)$
0	0	0	.10
0	0	1	.15
0	1	0	.10
0	1	1	.15
1	0	0	.25
1	0	1	0
1	1	0	.10
1	1	1	.15

Naive Bayes distribution:

A	B	C	$P(A\&B\&C)$
0	0	0	.127
0	0	1	.100
0	1	0	.072
0	1	1	.200
1	0	0	.222
1	0	1	.050
1	1	0	.127
1	1	1	.100

Example

Naive Bayes model learns wrong rules

- Classifies $A = 0, B = 0$ as $C = 0$
- Classifies $A = 1, B = 1$ as $C = 0$
- Best rule is to output $C = 1$ for these cases

Even though best rule is one it can represent!

Linear Classifiers

Equation (2D):

$$ax + by > 1$$

Or:

$$w_1x_1 + w_2x_2 > 1$$

Or (any number of dimensions):

$$w \cdot x > 1$$