Learning Naive Bayes Classifiers - An Example
Prof. Carolina Ruiz

Department of Computer Science WPI

Dataset:										
@relation credit-data										
<pre>@attribute credit_history {bad, unknown, good}</pre>										
@attribute debt {low, high}										
@attribute collateral {none, adequate}										
@attribute income {0-15, 15-35, >35}										
@attribute risk {low, moderate, high}										
@data	(,	, 8,								
bad,	low,	none,	0-15,	high						
unknown,	high,	none,	15-35,	high						
unknown,	low,	none,	15-35,	moderate						
bad,	low,	none,	15-35,	moderate						
unknown,	low,	adequate,	>35,	low						
unknown,	low,	none,	>35,	low						
unknown,	high,	none,	0-15,	high						
bad,	low,	adequate,	>35,	moderate						
good,	low,	none,	>35,	low						
good,	high,	adequate,	>35,	low						
good,	high,	none,	0-15,	high						
good,	high,	none,	15-35,	moderate						
good,	high,	none,	>35,	low						
bad,	high,	none,	15-35,	high						

Naïve Bayes Model Construction: Constructing a Naïve Bayes model over the above credit-data, where risk is the target attribute.

			RISK:		low modera		ate	high	
			probał	oility:	(5+1)/17	(4+1)/	(4+1)/17		
						_		I	
					Risk	_		COLL	ATERAL
	CR	EDIT-HIST	TORY	_ `	A		RISK	none	adequate
RISK	bad	unknown	good	_/			low	(3+1)/7	(2+1)/7
low	(0+1)/8	(2+1)/8	(3+1)/8		/ \ `		moderate	(3+1)/6	(1+1)/6
moderate	(2+1)/7	(1+1)/7	(1+1)/7		\backslash		high	(5+1)/7	(0+1)/7
high	(2+1)/8	(2+1)/8	(1+1)/8			7	$\overline{\ }$		
Credit Debt Collateral Income									>
-	DEBT CPT						INCOME		
-	RISK	low	high	-	RISK	0-15	15-3	5 >35	
	low	(3+1)/7	(2+1)/7		low	(0+1)/8	3 (0+1)/	/8 (5+1)/	/8
	moderate	(3+1)/6	(1+1)/6		moderate	e (0+1)/7	7 (3+1)/	/7 (1+1)/	7
	high	(1+1)/7	(4+1)/7		high	(3+1)/8	8 (2+1)/	/8 (0+1)/	/8

Classification using the Naïve Bayes Model:

Using the above Naïve Bayes model to classify a new instance:

Credit-History Debt Collateral Income Risk good low adequate 0-15 ?

predicted v = argmax P(Risk=v)

- * P(Credit-History=good | Risk=v)
- * P(Debt=low | Risk=v)
- * P(Collateral=adequate | Risk=v)
- * P(Income=0-15 | Risk=v)

v = low: P(Risk=low)

- * P(Credit-History=good | Risk=low)
- * P(Debt=low | Risk=low)
- * P(Collateral=adequate | Risk=low)
- * P(Income=0-15 | Risk=low)
- $= (6/17)^{*}(4/8)^{*}(4/7)^{*}(3/7)^{*}(1/8)$
- = 288/53312 = 0.0054

v = moderate: P(Risk=moderate)

- * P(Credit-History=good | Risk=moderate)
- * P(Debt=low | Risk=moderate)
- * P(Collateral=adequate | Risk=moderate)
- * P(Income=0-15 | Risk=moderate)
- $= (5/17)^{*}(2/7)^{*}(4/6)^{*}(2/6)^{*}(1/7)$
- = 80/29988 = 0.0027

v = high: P(Risk=high)

- * P(Credit-History=good | Risk=high)
- * P(Debt=low | Risk=high)
- * P(Collateral=adequate | Risk=high)
- * P(Income=0-15 | Risk=high)
- $= (6/17)^{*}(2/8)^{*}(2/7)^{*}(1/7)^{*}(4/8)$
- = 96/53312 = 0.0018

Hence, the predicted value by the Naïve Bayes classifier is Risk=low.