Learning Bayesian Nets from Data – An Example **Department of Computer Science**
Prof. Carolina Ruiz **WPI**

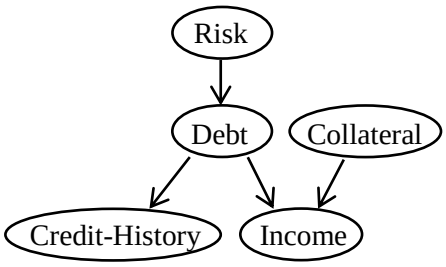
Dataset:

@relation credit-data
@attribute **credit_history** {bad, unknown, good}
@attribute **debt** {low, high}
@attribute **collateral** {none, adequate}
@attribute **income** {0-15, 15-35, >35}
@attribute **risk** {low, moderate, high}

@data

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

Bayes Net Construction:
Constructing a Bayes Net model over the above credit-data when the topology of the net is given:



We construct Probability Distribution Tables from the data:

RISK		
high	moderate	low
(5+1)/17	(4+1)/17	(5+1)/17

COLLATERAL	
none	adequate
(11+1)/16	(3+1)/16

	DEBT	
RISK	low	high
high	(1+1)/7	(4+1)/7
moderate	(3+1)/6	(1+1)/6
low	(3+1)/7	(2+1)/7

	CREDIT-HISTORY		
DEBT	bad	unknown	good
low	(3+1)/10	(3+1)/10	(1+1)/10
high	(1+1)/10	(2+1)/10	(4+1)/10

		INCOME		
DEBT	COLLATERAL	0-15	15-35	>35
low	none	(1+1)/8	(2+1)/8	(2+1)/8
low	adequate	(0+1)/5	(0+1)/5	(2+1)/5
high	none	(2+1)/9	(3+1)/9	(1+1)/9
high	adequate	(0+1)/4	(0+1)/4	(1+1)/4

Classification using the Bayesian Net: Using the above Bayesian Net to classify a new instance:

Credit-History	Debt	Collateral	Income	Risk
?	low	adequate	?	?

predicted classification= $\underset{v}{\text{argmax}}$ P(Risk=v | Debt=low & Collateral=adequate)

$$= \underset{v}{\text{argmax}} \frac{P(\text{Debt=low \& Collateral=adequate} \mid \text{Risk=v}) * P(\text{Risk=v})}{P(\text{Debt=low \& Collateral=adequate})}$$

$$= \underset{v}{\text{argmax}}\ P(\text{Debt=low \& Collateral=adequate} \mid \text{Risk=v}) * P(\text{Risk=v})$$

$$= \underset{v}{\text{argmax}}\ P(\text{Debt=low}|\text{Risk=v})*P(\text{Collateral=adequate}|\text{Risk=v})*P(\text{Risk=v})$$

$$= \underset{v}{\text{argmax}}\ P(\text{Debt=low}|\text{Risk=v})*P(\text{Collateral=adequate})*P(\text{Risk=v})$$

Now we calculate the values of the above expression for each possible value v of Risk:

v = low: P(Debt=low|Risk=v)*P(Collateral=adequate)*P(Risk=v)

= P(Debt=low|Risk=low)*P(Collateral=adequate)*P(Risk=low)

= (4/7)*(4/16)*(6/17)

= 0.0504

v = moderate: P(Debt=low|Risk=v)*P(Collateral=adequate)*P(Risk=v)

= P(Debt=low|Risk=moderate)*P(Collateral=adequate)*P(Risk=moderate)

= (4/6)*(4/16)*(5/17)

= 0.0490

v = high: P(Debt=low|Risk=v)*P(Collateral=adequate)*P(Risk=v)

= P(Debt=low|Risk=high)*P(Collateral=adequate)*P(Risk=high)

= (2/7)*(4/16)*(6/17)

= 0.0252

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

[Note that the value of P(Collateral=adequate) is independent of the what value v of Risk is under consideration. Hence, it would have sufficed above to use the expression P(Debt=low|Risk=v)*P(Risk=v) instead of P(Debt=low|Risk=v)*P(Collateral=adequate)*P(Risk=v) to decide what value v of Risk would be predicted by the Bayes Net Classifier.]
