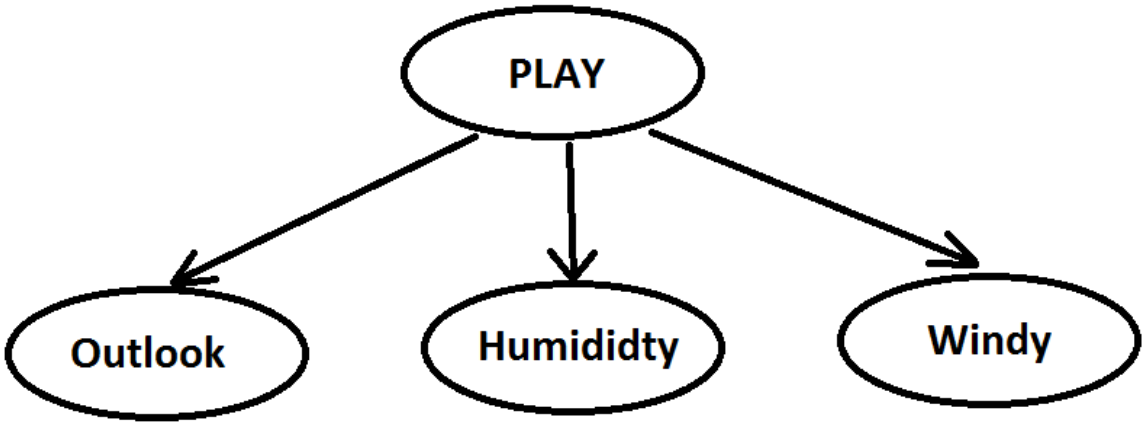
CS4445 Data Mining and Knowledge Discovery in Databases B Term 2014

Project 3. Part I. HOMEWORK ASSIGNMENT

Solutions by Artem Gritsenko

1. For the Naïve Bayes model we assume the conditional independence of the attributes given the target attribute. Thus, the network will look like:



Now, we need to calculate the Conditional Probability Tables for all the attributes. First, we compute the probability table for the target attribute Play:

Play	
Yes	**No**
(9+1)/16	(5+1)/16

Now, we compute the CPT tables for all other attributes.

Play	**Outlook**		
	Sunny (5)	**Overcast (4)**	**Rainy (5)**
Yes	(2+1)/12	(4+1)/12	(3+1)/12
No	(3+1)/8	(0+1)/8	(2+1)/8

Play	**Humidity**		
	Low (4)	**Normal (5)**	**High (5)**
Yes	(3+1)/12	(4+1)/12	(2+1)/12
No	(1+1)/8	(1+1)/8	(3+1)/8

Play	**Windy**	
	True (6)	**Overcast (8)**
Yes	(3+1)/11	(6+1)/11
No	(3+1)/7	(2+1)/7

2. We want to maximize the probability of the class value to be assigned to the data instance among all class values. Using the formulas for Naïve Bayes classification:

pred V =argmax(P(Play=v) * P(Outlook=Sunny|Play=v) * P(Humidity=normal|Play=v))*P(Windy|Play=v)

Since the value of Windy is not given, we need to consider each possible value of Windy:

P(Windy|Play=v) = P(Windy=True|Play=v) + P(Windy=False|Play=v) = 1. Hence,

pred V =argmax(P(Play=v) * P(Outlook=Sunny|Play=v) * P(Humidity=normal|Play=v)).

Now we use the above formula for each value of Play (Yes and No):

Pred "Play = Yes":

P(Play=Yes) * P(Outlook=Sunny|Play=Yes) * P(Humidity=normal|Play=Yes)

= 10/16 * 3/12 * 5/12 = 0.065

Pred "Play = No":

P(Play=No) * P(Outlook=Sunny|Play=No) * P(Humidity=normal|Play=No)

= 6/16 * 4/8 * 2/8 = 0.047

The probability for target value **Play = Yes** is higher than the probability for **Play = No**. Thus, the predicted value for the data instance's target attribute is **Yes**.

3. We build the CPT by taking into account all possible combinations of the values of attributes **Play** and **Humidity**, and counting the frequencies of the different values of **Outlook** for each case.

Play	**Humidity**	**Outlook**		
		Sunny	**Overcast**	**Rainy**
Yes	Low	(2+1)/6	(1+1)/6	(0+1)/6
Yes	Normal	(0+1)/7	(2+1)/7	(2+1)/7
Yes	High	(0+1)/5	(1+1)/5	(1+1)/5
No	Low	(0+1)/4	(0+1)/4	(1+1)/4
No	Normal	(1+1)/4	(0+1)/4	(0+1)/4
No	High	(2+1)/6	(0+1)/6	(1+1)/6