Exercise: 2

Meeting on January 12th / 14th

Problem 1: More on Naive Bayes

The following dataset is given:

Fall-Nr.	Clouds	Temperature	Humidity	Wind	Decision
1	sunny	high	high	no	+
2	sunny	high	high	yes	+
3	cloudy	medium	normal	no	+
4	rainy	high	high	no	-
5	sunny	low	normal	yes	-
6	rainy	low	high	yes	-
7	cloudy	high	high	yes	-
8	sunny	medium	normal	no	+
9	cloudy	medium	high	yes	-
Test case 1	rainy	medium	high	no	?
Test case 2	sunny	medium	high	yes	?
Test case 3	?	low	normal	yes	+
Test case 4	?	low	high	no	-

Table 1: Dataset and test cases for a Naive Bayes classifier

(a) Use Naive Bayes to predict the decisions for test cases 1 and 2 from the training data. What is the baseline for this problem?

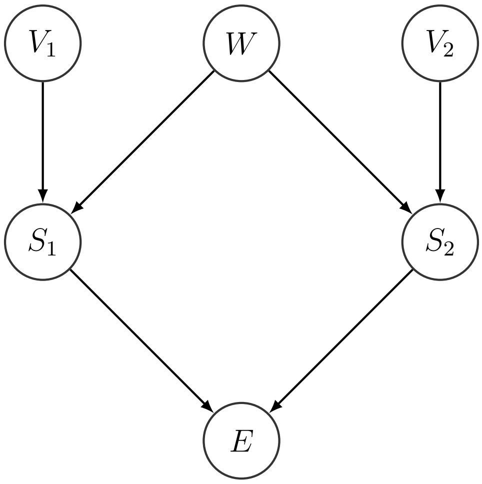
(b) Use Naive Bayes to predict the decisions for test cases 3 and 4 from the training data. What is the baseline for this problem?

(c) Draw the corresponding Bayes network for one of the two previous tasks and explain the assumptions made by the Naive Bayes model.

Problem 2: Tennis

Given is the following model for a game of tennis with two players (1 & 2):

Abbreviation	Description
V_1/V_2	Player 1/2 is hurt (v) or healthy (g)
W	The weather ist either sunny (s), cloudy (w) or rainy (r)
S_1/S_2	The player strength ot the two players
E	Outcome of the game: Either player 1 (e_1) or player 2 (e_2) wins.



a) What variable combinations need to be fulfilled so that S_1 and V_2 are d-separated? Make a table and write down for all possible combinations whether they fulfill the conditions.

b) Player strength is actually a continuous variable. Describe two approaches on how to model such a variable.

c) Assume that player strength has three possible values {strong, medium, weak}. How many entries does the conditional probablity table (CPT) for node S_1 contain? How many entries does the Bayes network contain in total? How many entries would the CPT need for variable E if all five remaining variables would have direct influence on the outcome E?

d) If S_1 contains the value "strong" and the weather is sunny, is V_1 dependent on the other variables? Explain your answer.

e) Now assume the following value ranges and CPTs. What is the prior probability for the weather being cloudy or rainy? What is the probability if you know that both players are "healthy", but have a "weak" player strength?

V_1	P
hurt v_1	0,3
healthy g_1	0,7

V_2	P
hurt v_2	0,2
healthy g_2	0,8

W	P
sunny s	0,6667
cloudy w	0,2222
rainy r	0,1111

Weather W is	sunny s		cloudy b		rainy r	
Health V_1 is	hurt	healthy	hurt	healthy	hurt	healthy
Player 1 is strong $\uparrow_1$	0,2	0,6	0,5	0,6	0,6	0,7
Player 1 ist weak $\downarrow_1$	0,8	0,4	0,5	0,4	0,4	0,3

Weather W is	sunny s		cloudy b		rainy r	
Health V_1 is	hurt	healthy	hurt	healthy	hurt	healthy
Player 2 is strong $\uparrow_2$	0,6667	0,6667	0,3333	0,3333	0,25	0,25
Player 2 is weak $\downarrow_2$	0,3333	0,3333	0,6667	0,6667	0,75	0,75

Player 1 is	strong		weak	
Player 2 is	strong	weak	strong	weak
Player 1 wins e_1	0,4	0,2	0,6	0,5
Player 2 wins e_2	0,6	0,8	0,4	0,5

Problem 3: Sampling-Methods

In the year 2031 you are leading a company making skirts for people with plus size worn by both men and women. You deliver to England (80%) and Scotland (20%). From your statistical analysis you know that both countries have less women f than men m with a ratio of 4/6. The probability of a person wearing skirts as well as the probability of a person needing plus size are given in 2.

Country	Probability
England	80%
Scotland	20%

Gender	Probability
Female	40%
Male	60%

	Scotland		England	
	male	female	male	female
Plus size	16%	3%	50%	6%
Skirt	5%	30%	0,5%	40%

Table 2: Statistics about skirt size and wearing skirts.

Now view the following Bayes network in Fig. 1. You know from your statistics that

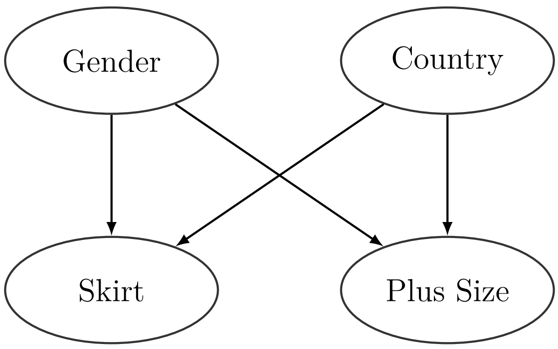


Figure 1: Bayes Network

60 million people are wearing plus size. Now you want to know the size of your target group.

(a) What are the two possibilites to compute $P(\text{Wearing Skirts}|\text{Plus Size})$?

(b) You are tired of analytical methods and instead remember a special Markov-Chain-Monte-Carlo (MCMC) sampling method called Gibbs sampling. Describe the method and its differences to Likelihood-Weighting and direct sampling.

(c) You know about tools like UnBBayes[1], which implement these algorithms but as a computer science graduate you would like to refresh your knowledge about how such an algorithm works. For precise inspection you choose the Likelihood-Weighting-Method. Complete 3 with the missing samples.

Node: Name	Gender	Country	Plus Size	Skirt	Weight
Next random number	0.60	0.91	-	0.08	
Node: Value	m	S	yes	no	0.16
Next random number	0.55	0.04	-	0.21	
Node: Value	m	E	yes	no	0.50
Next random number	0.80	0.79	-	0.58	
Node: Value					
Next random number	0.91	0.85	-	0.34	
Node: Value					
Next random number	0.86	0.97	-	0.02	
Node: Value					
Next random number	0.21	0.26	-	0.06	
Node: Value					
Next random number	0.17	0.60	-	0.88	
Node: Value					
Next random number	0.60	0.66	-	0.13	
Node: Value					
Next random number	0.70	0.67	-	0.68	
Node: Value					
Next random number	0.39	0.71	-	0.37	
Node: Value					

Table 3: Samples for Liklihood-Weighting

(d) Now you want to estimate $P(\text{Wearing Skirts}|\text{Plus Size})$ from these samples. Use

$$P(A|B) = \frac{\sum \text{Weights of samples that fulfill condition(s) } A}{\sum \text{Weights of all samples}}.$$

What is the size of your target group?

(e) Why is $P(\text{Wearing Skirts}|\text{Plus Size})$ not exact? How can you improve the accuracy?

(f) Implement the algorithm for the example above and compute a better value for $P(\text{Wearing Skirts}|\text{Plus Size})$. If necessary, plot multiple curves with different starting conditions against the number of samples.

[1] http://unbbayes.sourceforge.net/

Problem 4: Part-of-speech

One of the many practical application of (dynamic) Bayes networks is Part-of-Speech tagging, i.e. assigning parts of speech to each word in a text. Similar to the umbrella example from the lecture, Part-of-Speech taggers model words as evidence variables and the corresponding part-of-speech tag as hidden states.

In the following example we assume a very simplified tag-set given in Tab. 4).

Abbreviation	Description	Examples
DET	determiner	the, a
N	noun	year, home, costs, time, education
PRO	pronoun	he, their, her, its, my, I, us
V	verb	said, took, told, made, asked

Table 4: Simplified tag-set.

For instance in the case of "I like you" a possible assignment of variables of the net in Fig. 2 would be given by t_1 = PRO, w_1 = I, t_2 = V, w_2 = like, t_3 = PRO and w_3 = you. Every t_i is a part-of-speech tag and every w_i is a word.

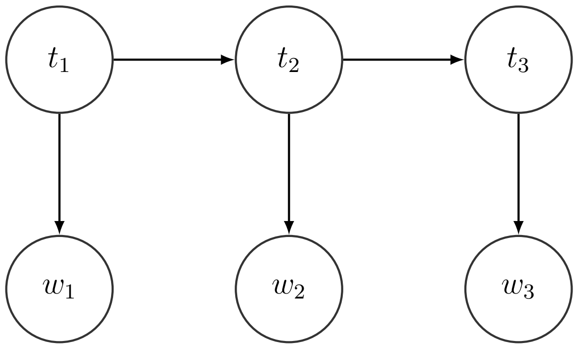


Figure 2: Beispiel zum Part-of-Speech Modell

The part-of-speech transition probabilities $P(t_i|t_{i-1})$ are given in Tab. 5 and an excerpt of word emission probabilities $P(w_i|t_i)$ is given in Tab. 6.

from/to	DET	N	PRO	V
DET	0,05	0,90	0,05	0
N	0,10	0,70	0,05	0,15
PRO	0,05	0,50	0,05	0,40
V	0,50	0,10	0,40	0

Table 5: Part-of-speech transition probabilities

For simplicity, for the beginning of the sentence (first tag) you can assume a uniform distribution of tags.

(a) Compute the most likely sequence of tags of the sentence: "I saw the saw". Use the Viterbi algorithm.

(b) What factors influence the complexity of the computation and how are they connected?

w/t	DET	N	PRO	V
I	0	$2 \cdot 10^{-4}$	0,1	0
saw	0	$8 \cdot 10^{-4}$	0	0,1
the	0,1	0	0	0

Table 6: Excerpt of word emission probabilities