P(B | WB) = P(WB | B) * P(B) / P(WB) = 0.75 * 0.8 / P(WB)
P(~B | WB) = P(WB |~B) * P(~B) / P(WB) = 0.25 * 0.2 / P(WB)
P(B|WB) + P(~B|WB) = 1
From:
0.75 * 0.8 / P(WB) + 0.25 * 0.2 / P(WB) = 1
0.6 / P(WB) + 0.05/P(WB) = 1
it follows that: P(WB) = 0.65
Thus P(B|WB) = 0.75 * (0.8/0.65) = 0.923

P(B|WB1,WB2) = P(WB1,WB2|B) * (P(B)/P(WB1,WB2))
= P(WB1|B)*(P(WB2|B)*(P(B)/P(WB1,WB2))
= 0.75*0.75*(0.8/P(WB1,WB2))

P(~B | WB1,WB2) = P(WB1 |~B) * P(WB2|~B) * P(~B) / P(WB)
= 0.25 * 0.25 * 0.2 / P(WB1,WB2)

P(B|WB1,WB2) + P(~B|WB1,WB2) = 1

From:
0.75 * 0.75 * 0.8 / P(WB1,WB2) + 0.25 * 0.25 * 0.2 / P(WB1,WB2) = 1
0.45 / P(WB1,WB2) + 0.0125 / P(WB1,WB2) = 1
it follows that: P(WB1,WB2) = 0.4625= 37/80

(0.75*0.75*0.8)/(37/80)

P(B|WB1,WB2) = (0.75*0.75*0.8)/(37/80) = 0.973

The chances that the car is behind doors D_1, D_2, D_3 are equal: 1/3. Let O_j be "The quiz master opens D_j. Let K_j be the candidate selects first door D_j. Assume K_1 and that the quiz master does O_3. Then: $P(O_3|D_3) = 0$, $P(O_3|D_2) = 1$ en $P(O_3|D_1) = 1/2$.
According the law of total probability (wet van de totale kans):

$$P(O_3) = P(O_3|D_1).P(D_1) + P(O_3|D_2).P(D_2) + P(O_3|D_3).P(D_3)$$
$$= 1/2 * 1/3 + 1 * 1/3 + 0 * 1/3 = 1/2$$

$$P(D_1|O_3) = \frac{P(O_3|D_1)P(D_1)}{P(O_3)} = \frac{1/2 * 1/3}{1/2} = 1/3.$$

Similarly: $P(D_2|O_3) = 2/3$. Hence, it is better to switch after the door is opened.

A bag H_1 contains 10 marbles: 2 red, 3 white and 5 blue. Bag H_2 also has 10 marbles, now: 3 red, 3 white and 4 blue. John takes 5 times a marble from one of the bags. After each take he puts the marble back into the bag before he does the next take.

1. What is the probability of outcome $D =< 2,1,2 >$ (i.e. 2 red, 1 white and 2 times blue) if he takes from bag H_1?
2. What is the probability of $D =< 2,1,2 >$ if he takes them from H_2?
3. Suppose he has chosen randomly one of the two bags before he takes 5 times a marble out of the chosen bag. The outcome is as above: 2 red, 1 white en 2 times blue. From which bag were the marbles taken?
4. Suppose he chooses the bag not randomly but after a throw with a die. When the outcome is 3 or 6 he takes bag H_2, otherwise he takes H_1. What is the most likely bag chosen when the outcome is $D =< 2,1,2 >$?

1. Compute $P(< 2,1,2 > |H_1)$ using the formula for the multinomial distribution $Mult(5, < 0.2, 0.3, 0.5 >)$.
   $P(< 2,1,2 > |H_1) = 30 * (0.2)^2 * (0.3)^1 * (0.5)^2 = 0.09$
2. Compute $P(< 2,1,2 > |H_2)$ using the formula for the multinomial distribution $Mult(5, < 0.3, 0.3, 0.4 >)$.
   $P(< 2,1,2 > |H_2) = 30 * (0.3)^2 * (0.3)^1 * (0.4)^2 = 0.1296$
3. Given the outcome both bags are possible. We compute $P(H_1| < 2.1.2 >)$ en $P(H_2| < 2,1,2 >)$ with Bayes' Rule.
   $P(H_1| < 2,1,2 >) = P(< 2,1,2 > |H_1) * P(H_1)/P(< 2,1,2 >) = 0.09 * 0.5/P(< 2,1,2,>)$
   According the law of total probability (marginalisation):
   $P(< 2,1,2 >) = P(< 2,1,2 > |H_1) * P(H_1) + P(< 2,1,2 > |H_2) * P(H_2) = 0.09 * 0.5 + 0.1296 * 0.5 = 0.1098$
   Hence: $P(H_1| < 2,1,2,>) = 0.045/0.1098 = 0.4098 \approx 0.41$
   Similarly:
   $P(H_2| < 2,1,2 >) = P(< 2,1,2 > |H_2) * P(H_2)/P(< 2,1,2 >) = 0.1296 * 0.5/P(< 2,1,2,>)$
   Hence: $P(H_2| < 2,1,2,>) = 0.0648/0.1098 \approx 0.59$
   Conclusion: H_2 is the most likely bag.
   Note that in order to compute the most likely bag, one does not need to compute the full terms. Constant factors can be removed.
4. Similar as the previous question but now the priors are $P(H_1) = 2/3$ and $P(H_2) = 1/3$ instead of 0.5 as in the previous case.

Name	Tall	Eye Color	Hair Length	Sex
Drew	No	Blue	Short	Male
Claudio	Yes	Brown	Long	Female
Drew	No	Blue	Long	Female
Drew	No	Blue	Long	Female
Alberto	Yes	Brown	Short	Male
Karin	No	Blue	Long	Female
Nina	Yes	Brown	Short	Female
Sergio	Yes	Blue	Long	Male

A person called Drew is blue-eyed, tall, and has long hair. Use Naive Bayes to predict the most likely gender of this person.

$$P(Drew|Female) \times P(Blue|Female) \times P(Tall|Female) \times P(LongHair|Female) \times P(Female)$$

$$= 2/5 \times 3/5 \times 2/5 \times 4/5 \times 5/8 = 0.048$$

with:

$$P(Drew|Male) \times P(Blue|Male) \times P(Tall|Male) \times P(LongHair|Male) \times P(Male)$$

$$= 1/3 \times 2/3 \times 2/3 \times 1/3 \times 3/8 = 0.0185$$

2 *(20 points)* Given the following piece of text from an email:

attention if you are in debt. if you are then we can help. qualifying is now at your fingertips and there are no long distance calls

(a) Assume that we use as vocabulary $V =$ {attention, adult, debt, publications, qualifying, xxx }. How would this piece of text be coded using a binary coding and this vocabulary V?

(b) For convenience consider a smaller vocabulary $V =$ {attention, adult, debt} and assume that we have a dataset consisting of 100 emails of which 30 are spam and with the following vocabulary frequency list:

Word	Ham	Spam
attention	30	10
adult	0	22
debt	4	20

This means for instance that the word "attention" occurs in 30 ham emails and in 10 spam emails. Assume that a new email arrives with binary coding $< 1,0,1 >$. Compute the likelihood that this email is from the spam class. in other words compute $P(< 1,0,1 > |Spam)$.

(c) How is this new email with coding $< 1,0,1 >$ classified; *Ham* or *Spam*, if one uses a Naive Bayes approach with no smoothing?

Antwoord op 2.

1. $< 1,0,1,0,1,0 >$
2. $P(< 1,0,1 > |Spam) = \frac{10}{30}\frac{8}{30}\frac{20}{30} = \frac{16}{270} = 0.0593$
3. $P(< 1,0,1 > |Ham) = \frac{12}{490} = 0.0245.$ (=30/70*70/70*4/70)
4. It is easily computed that $P(< 1,0,1 > |Ham)\frac{7}{10}/P(< 1,0,1 > |Spam)\frac{3}{10} = 0.964 < 1$. Hence email is classified as Spam.

High recall: Everything we should filter will be filtered.

High precision: Everything we have filtered was supposed to be filtered.

the caveats are as following:

High recall: We might filter more than we should.

High precision: We might not filter everything that should be filtered.

The relation between a line $w_0 + w_1x_1 + w_2x_2 = 0$ and a classification is very simple. All points (x_1, x_2) in the two dimensional feature space will be classified as 1 if $w_0 + w_1x_1 + w_2x_2 > 0$ and classified as 0 otherwise.

$$(w_0^{new}, w_1^{new}, w_2^{new}) = (w_0^{old}, w_1^{old}, w_2^{old}) - \alpha(1, b_1, b_2)$$

Classified as 1 instead of 0. Use + if coordinate b is classified as 0 but is 1

Accuracy: the fraction of correctly classified examples
Error rate: the fraction of misclassified examples

		Predicted class	
		C_1	C_2
Actual Class	C_1	a	b
	C_2	c	d

$$Accuracy = \frac{a+d}{a+b+c+d}$$
$$Error\ rate = \frac{b+c}{a+b+c+d}$$
$$Recall\ for\ class\ C_1 = \frac{a}{a+b}$$
$$Precision\ for\ class\ C_1 = \frac{a}{a+c}$$

Entropy (a statistical property that measures how dispersed or scattered a data set is):
A pure set: 0
A 50% 50% set: 1

$$E(D) = \sum_{i=1}^{i=c} -p_i \log_2(p_i)$$

Tree: Calculate infoGain for all columns; Highest column is the root; Do the same but now for specified rows

Feature which reduces the entropy the most if the examples were to be partitioned according to that feature:

$$Gain(D, A) = E(D) - \sum_{j=1}^{j=v} \frac{|D_j|}{|D|} E(D_j)$$

Example: 9 positive assesments, 11 negative.
E(D) = -9/20 $\log_2$(9/20) – 11/20$\log_2$(11/20)
= 0.9928 (initial entropy)

A, G, P, The set of examples for which T has the value A, called D_A, consists of 6 examples $\{3, 6, 8, 11, 15, 17\}$ of which two have a positive assessment. Hence $E(D_A) = -2/6 \log_2(2/6) - 4/6 \log_2(4/6) = 0.9183$.

Calculate the other two values the same way:

Hence the information gain for feature T is given by $Gain(D, T) = 0.9928 - 6/20 * 0.9183 - 9/20 * 0.7642 - 5/20 * 0 = 0.3734$.

Compute the confidence we have that A is true given new data B:

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A)$$

Exercise 2. A bag contains 5 fair dice: a 4 sided die, a 6 sided one, an 8, a 12 and a 20 sided die. They are all fair dice, so that for example the probability of throwing a 3 with a 20 sided die is 1/20. John randomly draws a die from the bag. What is the prior probability that John draw the 4 sided die? That is 1/5, the distribution of the probability of taking each of the dice is uniform. John throws with the die he has drawn from the bag. The outcome is 5. This is the data D. What is the most likely value of H when we know that D (i.e. that the outcome is 5)? Use Bayes' belief update rule.

H	$P(H)$	$P(D\|H)$	$P(D\|H).P(H)$	$P(H\|D)$
4	1/5	0	0	0
6	1/5	1/6	1/30	0.392
8	1/5	1/8	1/40	0.294
12	1/5	1/12	1/60	0.196
20	1/5	1/20	1/100	0.118
	1	x	0.085	1

H	P(H)	P(D|H)	P(D|H).P(H)	P(H|D)
4	1/5	1/4	1/20	0,370
6	1/5	1/6	1/30	0,247
8	1/5	1/8	1/40	0,185
12	1/5	1/12	1/60	0,123
20	1/5	1/20	1/100	0,074
	1	27/40 0,675	27/200 0,135	1

H	P(H|D=5)	P(H|D=1)	P(H|D=5).P(H|D=1)
4	0	0,370	0
6	0,392	0,247	0,097
8	0,294	0,185	0,054
12	0,196	0,123	0,024
20	0,118	0,074	0,009

4 *(20 points)* A certain classifier was tested on a test, resulting in the following confusion matrix:

		Predicted class		
		C_1	C_2	C_3
Actual Class	C_1	120	15	20
	C_2	16	150	10
	C_3	22	3	130

(a) What is the accuracy of this classifier?
(b) What is the recall of this classifier for class C_2?
(c) What is the precision of this classifier for class C_3?

Antwoord op 4.

1. Accuracy: $[120+150+130]/[120+150+130+15+20+10+16+22+3] = 0.82$
2. Recall C_2: $150/[16+150+10] = 0.85$
3. Precision C_3: $130/[130+10+20] = 0.81$