CMPT726 - Machine Learning
Assignment 1

Topic: Probabilistic Reasoning, Maximum Likelihood, Classification
Submitted to: Professor Oliver Schulte
Submitted by: Raquel Yuri da Silveira Aoki
Date: October 10, 2017

Question 1 - Joint and Conditional Probabilities

1) Use the product formula of Bayes nets and the conditional probability parameters specified by AIspace to compute the probability that: all nodes are true.

Solution
Considering $Influenza = I, Smokes = S, SoreThroat = ST, Fever = F, Bronchitis = B, Coughing = C$ and $Wheezing = W$ and using proprieties of bayesian networks:

$P(W = T, C = T, B = T, S = T, F = T, I = T, ST = T) =$
$P(W = T|B = T)P(C = T|B = T)P(B = T|S = T, I = T)P(F = T|I = T)P(ST = T|I = T)P(S = T)P(I = T) = 0.6 \times 0.8 \times 0.99 \times 0.9 \times 0.3 \times 0.05 \times 0.2 = 0.00128304$

2) Use the product formula of Bayes nets and the conditional probability parameters specified by Aispace to compute the probability that: all nodes are true except for Sore Throat, and that Sore Throat is false.

Solution
$P(W = T, C = T, B = T, S = T, F = T, I = T, ST = F) =$
$P(W = T|B = T)P(C = T|B = T)P(B = T|S = T, I = T)P(F = T|I = T)P(ST = F|I = T)P(S = T)P(I = T) = 0.6 \times 0.8 \times 0.99 \times 0.9 \times 0.7 \times 0.05 \times 0.2 = 0.00299376$

3) Show how can you use these two joint probabilities to compute the probability that: all nodes other than Sore Throat are true. (Where the value of Sore Throat is unspecified.)

Solution
Using the property $P(A) = P(A, B) + P(A, \neg B)$ in this context, we have

$P(W=T, C=T, B=T, S=T, F=T, I=T) =$
$P(W=T, C=T, B=T, S=T, F=T, I=T, ST=T) + P(W=T, C=T, B=T, S=T, F=T, I=T, ST=F) = 0.00128304 + 0.00299376 = 0.0042768$

4) Verify the product formula: $P(allnodesaretrue) = P(SoreThroat = true|allothernodesaretrue)xP(allothernodesaretrue)$. You may get the first conditional probability by executing a query with the tool.

Solution
Using $P(A,B) = P(A|B)P(B) = \frac{P(A,B)}{P(B)}P(B) = P(A)$, if $P(B) \neq 0$
in this context, we have:
$P(W=T, C=T, B=T, S=T, F=T, I=T, ST=T) = P(W, C, B, S, F, I, ST) =$
$P(ST|W, C, B, S, F, I) \times P(W, C, B, S, F, I) = \frac{P(W,C,B,S,F,I,ST)}{P(W,C,B,S,F,I)} \times P(W, C, B, S, F, I) =$
$P(W, C, B, S, F, I, ST)$ if $P(W, C, B, S, F, I) \neq 0$

5) Compute the probability that Sore Throat is true and that Fever is true. (Hint: If you use the right formula, you need only 4 conditional probabilities.)

Solution
$P(ST=T, F=T) = P(ST, F) = P(ST, F|I)P(I) + P(ST, F|I=F)P(I=F) =$
$P(ST|I)P(F|I)P(I) + P(ST|I=F)P(F|I=F)P(I=F) =$
$0.3 \times 0.9 \times 0.05 + 0.001 \times 0.05 \times 0.95 = 0.0135475$

Table 1: Summary Table - Question 1

Probability to be Computed	Your Result
P(all nodes true)	0.00128304
P(Sore Throat = False, all other nodes true)	0.00299376
P(all nodes other than Sore Throat true)	0.0042768
P(all nodes are true) = P(Sore Throat = true | all other nodes are true) x P(all other nodes are true)	NA
P(Sore Throat = true, Fever = True)	0.0135475

Question 2 - Theory: Conditional Probabilities

Exercise 13.3 in Russell and Norvig AMAI
For each of the following statements, either prove it is true or give a counter example.

a) If $P(a|b,c) = P(b|a,c)$, then $P(a|c) = P(b|c)$

Solution: True
Expanding the statement $P(a|b,c) = P(b|a,c)$:
$P(a|b,c) = \frac{P(a,b,c)}{P(b,c)} = P(b|a,c) = \frac{P(a,b,c)}{P(a,c)}$
Thus: $\frac{P(a,b,c)}{P(b,c)} = \frac{P(a,b,c)}{P(a,c)}$ and $P(b,c) = P(a,c)$.
Dividing both sides by $P(c)$: $P(b,c)/P(c) = P(a,c)/P(c) \Rightarrow$P(a|c) = P(b|c).

b)If $P(a|b,c) = P(a)$, then $P(b|c) = P(b)$

Solution: False
Consider the follow events:
a: *select randomly the king of hearts in a 52-card deck*
b: *select randomly a number $\leq$ 3 in a dice*
c: *select randomly a odd number in the same dice of b.*
The probabilities of those events are:
$P(a = true) = 1/52$, $P(b) = 1/2$,$P(c) = 1/2$,$P(b,c) = \frac{\#\{1,3\}}{\#\{1,2,3,4,5,6\}} = 2/6$
So:
$P(a|b,c) = \frac{P(a,b,c)}{P(b,c)} = \frac{\frac{1}{52}\times\frac{2}{6}}{\frac{2}{6}} = \frac{1}{52} = P(a)$
However:
$P(b|c) = \frac{P(b,c)}{P(c)} = \frac{2/6}{1/2} = \frac{4}{6} \neq P(b) = \frac{1}{2}$
which is a counter example of the statement.

c) If $P(a|b) = P(a)$, then $P(a|b,c) = P(a|c)$

Solution: False
Considering just the nodes Influenza, Smokes and Bronchitis from *Simple Diagnostic Example* as counter example: If $a = smokes$, $b = influenza$ and $c = Bronchitis$, then:
$P(a = t|b = t) = 0.2 = P(a = t)$
But:
$P(a = t|b = t, c = t) = \frac{P(a,b,c)}{P(b,c)} = \frac{0.0099}{0.0459} = 0.2156 \neq P(a|c) = \frac{0.1439}{0.1789} = 0.8043$
showing that the statement is false.

Question 3 - Theory Expectations

Show that $E[X_1 + X_2] = E[X_1] + E[X_2]$:

Solution

Expectation definition in the discrete case: $E[X] = \sum_x (x \times p(x))$

Considering $f(X_1, X_2, X_3) = X_1+X_2+X_3$, the expectation of $E[f(X_1, X_2, X_3)]$ can be obtained as follows::

$$
\begin{aligned}
E[X_1 + X_2 + X_3] &= \sum_{x_1,x_2,x_3} ((X_1 + X_2 + X_3) \times P(X_1, X_2, X_3)) \\
E[X_1 + X_2 + X_3] &= \sum_{x_1,x_2,x_3} (X_1 \times p(X_1, X_2, X_3)) + \sum_{x_1,x_2,x_3} (X_2 \times p(X_1, X_2, X_3)) + \\
&\quad \sum_{x_1,x_2,x_3} (X_3 \times p(X_1, X_2, X_3))
\end{aligned}
\qquad \text{(Eq. 1)}
$$

Focusing on the first term of the eq. (1) $\sum_{x_1,x_2,x_3}(X_1 \times p(X_1, X_2, X_3))$, we have that $\sum_{x_1,x_2,x_3}(X_1 \times p(X_1, X_2, X_3)) = \sum_{x_1}(X_1 \times p(X_1))$. This propriety occur because it was made a sum over all elements of $P(X_2)$ and $P(X_3)$, thus: $\sum_{x_2} p(X_1, X_2, X_3) = 1$ and $\sum_{x_3} p(X_1, X_2, X_3) = 1$. Applying the same argument for the others parts of eq. (1):

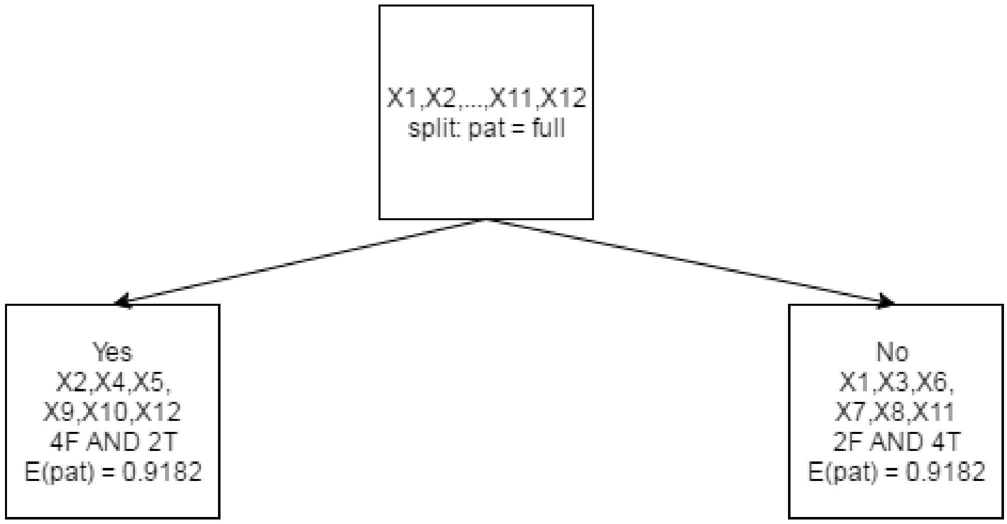
$$
\begin{aligned}
E[X_1 + X_2 + X_3] =& \sum_{x_1}(X_1 \times p(X_1)) \times 1 \times 1 + \\
& \sum_{x_2}(X_2 \times p(X_2)) \times 1 \times 1 + \\
& \sum_{x_3}(X_3 \times p(X_3)) \times 1 \times 1 \\
E[X_1 + X_2 + X_3] =& E[X_1] + E[X_2] + E[X_3]
\end{aligned}
$$

Question 4 - Practice: Decision Tree Learning with ID3 (10 marks)

Figure 1 provides data about whether a customer will wait for a table in a restaurant or not. Assume that ID3 splits first on the Pat attribute (for Patrons). Show the following for the branch $P\ at = \text{Full}$.

1) The next attribute chosen by ID3. There may be a tie among several attributes; you can list all or just one of them.

Solution



Varible	Entropy	Gain	New Groups
Alt= T	0.8091	0.1091	F, F
Bar = T	0.9182	0.0000	F, F
Hun = T	**0.6666**	**0.2516**	**F, tie**
Price = 1	**0.6666**	**0.2516**	**F, tie**
Rain = T	0.8091	0.1091	F. F
Res = T	**0.6666**	**0.2516**	**F, tie**
Type = F or I	**0.6666**	**0.2516**	**F, tie**
Est > 60	**0.6666**	**0.2516**	**F, tie**

Varible	Entropy	Gain	New Groups
Alt= T	0.8091	0.1091	T, T
Bar = T	0.9182	0.0000	T, T
Hun = T	**0.4591**	**0.4591**	**T, F**
Price = 1	**0.4591**	**0.4591**	**F, T**
Rain = T	0.9182	0.0000	T, T
Res = T	**0.4591**	**0.4591**	**T, F**
Type = F or I	0.4591	**0.4591**	T, TIE

Figure 1: Question 4.1 - Decision Tree

Considering the exercise, it was asked to show which is the next attribute chosen by ID3 in $pat = full$). The best splits in this node are Hun, Price, Res, Type and Est with the Gain of 0.2516. However, none of those attributes are truly good once that they will divide the items in one group 'Will Wait = F' and a second group with a tie between 'WillWait = F' and 'WillWait = T'.

As the Figure 1 shows, the best options for the next split are the variables Hun, Price or Res on the right child node (parent node $pat! = full$). They are the best options because they have the highest gain and one of their new child nodes will be $WillWait = T$ and the other $WillWait = F$.

2) The expected information gain associated with the next attribute. Compare this with the expected information gain for Hungry.

Solution

Hungry in the right child node is one of the variables with max information gain associated. The in left child node it also has a gain bigger than other variables, but its child nodes will not separate well the elements, once that one group it will be F and in the other it will have a tie between F and T.

3) How you calculated the expected information gain

Solution

1. Parent node entropy: $E(pat) = -(4/6) \times log_2(4/6) - (2/6) \times log_2(2/6) = 0.9182958$
2. For each possible split, it was calculated $E(variable, pat)$. For example, for the variable Hungry in the child node where $Pat = Full$ is false: $E(Hungry, Pat) = (3/6) \times (-(3/3) \times log_2(3/3)) + (3/6) \times (-(2/3) \times log_2(2/3) - (1/3) \times log_2(1/3)) = 0.459147917$.
3. Thus, the Gain is $E(pat) - E(Hungry, Pat) = 0.9182958 - 0.459147917 = 0.4591479$

Question 5 - Maximum Likelihood Parameter Estimation for Bayesian Networks (8 marks)

1) Write down the likelihood and the log-likelihood of the training data given a parameter setting of the Bayes net. Please use the notation.

Solution

$$f(x|\theta) = \prod_{n=1}^{N} P(X_1)P(X_2)P(X_3)P(X_4|X_1, X_2, X_3, X_4)$$
$$\prod_{k=1}^{L}(\theta_{10k}^{n_{10k}}) \prod_{k=1}^{L}(\theta_{20k}^{n_{20k}}) \prod_{k=1}^{L}(\theta_{30k}^{n_{30k}}) \prod_{j=1}^{L^3} \prod_{k=1}^{L}(\theta_{4jk}^{n_{1jk}})$$

Thus, the log-likelihood will be:

$$log(f(x|\theta)) = \sum_{k=1}^{L} n_{10k} ln(\theta_{10k}) + \sum_{k=1}^{L} n_{20k} ln(\theta_{20k}) + \\ \sum_{k=1}^{L} n_{30k} ln(\theta_{30k}) + \sum_{j=1}^{L^3} \sum_{k=1}^{L} n_{4jk} ln(\theta_{4jk})$$

2) Show that with binary nodes (L = 2), the maximum likelihood parameter values $\hat{\theta}_{ijk}$ are the conditional frequencies observed in the data.
Solution

Changing the equation for L = 2:

$$log(f(x|\theta)) = \sum_{k=1}^{2} n_{10k} ln(\theta_{10k}) + \sum_{k=1}^{2} n_{20k} ln(\theta_{20k}) + \\ \sum_{k=1}^{2} n_{30k} ln(\theta_{30k}) + \sum_{j=1}^{8} \sum_{k=1}^{2} n_{4jk} ln(\theta_{4jk})$$

Where each term can be express as:

$$\sum_{k=1}^{2} n_{i0k} ln(\theta_{i0k}) = n_{i01} ln(\theta_{i01}) + n_{i02} ln(1 - \theta_{i01})$$

Taking as example $i = 1$:

$$\begin{aligned} \frac{\partial log(f(x|\theta))}{\partial \theta_{101}} =& n_{101} \frac{1}{\theta_{101}} + n_{102} \frac{1}{1 - \theta_{101}} (-1) + 0 + 0 + 0 = 0 \\ & n_{101} \frac{1}{\hat{\theta}_{101}} - n_{102} \frac{1}{1 - \hat{\theta}_{101}} = 0 \\ & n_{101}(1 - \hat{\theta}_{101}) - n_{102} \hat{\theta}_{101} = 0 \\ & \hat{\theta}_{101}(n_{101} + n_{102}) = n_{101} \\ & \hat{\theta}_{101}(\sum_{k'} n_{10k'}) = n_{101} \end{aligned}$$

$$\hat{\theta}_{101} = \frac{n_{101}}{\sum_{k'} n_{10k'}}$$

The same argument can be use to i = 1, 2 and 3; and for k = 1 or 2. Thus:

$$\hat{\theta}_{i0k} = \frac{n_{i0k}}{\sum_{k'} n_{i0k'}}, \forall i = 1, 2, 3; k = 1, 2$$

For $i = 4$:

$$\begin{aligned}
\sum_{j=1}^{8}\sum_{k=1}^{2} n_{4jk}ln(\theta_{4jk}) + \sum_{j=1}^{8}(n_{4j1}ln(\theta_{4j1}) + n_{4j2}ln(1-\theta_{4j1}))\\
&= (n_{411}ln(\theta_{411}) + n_{412}ln(1-\theta_{411}))\\
&+ (n_{421}ln(\theta_{421}) + n_{422}ln(1-\theta_{421}))\\
&+ (n_{431}ln(\theta_{431}) + n_{432}ln(1-\theta_{431}))\\
&+ (n_{441}ln(\theta_{441}) + n_{442}ln(1-\theta_{441}))\\
&+ (n_{451}ln(\theta_{451}) + n_{452}ln(1-\theta_{451}))\\
&+ (n_{461}ln(\theta_{461}) + n_{462}ln(1-\theta_{461}))\\
&+ (n_{471}ln(\theta_{471}) + n_{472}ln(1-\theta_{471}))\\
&+ (n_{481}ln(\theta_{481}) + n_{482}ln(1-\theta_{481}))
\end{aligned}$$

And the same previous argument can be used to proof that $\hat{\theta}_{ijk} = \frac{n_{ijk}}{\sum_{k'} n_{ijk'}}, \forall i = 4; k = 1, 2; j = 1, 2, ..., 8$. Thus, the argument is valid to all i and j in binary problems.

Question 6 - Maximum Likelihood Parameter Estimation for a Gaussian Distribution (12 marks)

1) Write dow n the log-likelihood function

Solution

$$\begin{aligned}
L(\mathbf{x}, \mu, \sigma^2) &= \prod_{i=1}^{N}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} \times exp\{-\frac{(x^i-\mu)^2}{2\sigma^2}\}\right\}\\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \times exp\{-\frac{\sum_{i=1}^{N}(x^i-\mu)^2}{2\sigma^2}\}\\
ln(L(\mathbf{x}, \mu, \sigma^2)) &= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{\sum_{i=1}^{N}(x^i-\mu)^2}{2\sigma^2}
\end{aligned}$$

2) Show that the maximum likelihood estimate for the distribution mean

Solution

$$\hat{\mu} = argmax(\frac{\partial ln(L(\mathbf{x}, \mu, \sigma^2))}{\partial \mu})$$

$$\frac{\partial ln(L(\mathbf{x}, \mu, \sigma^2))}{\partial \mu} = 0 - 0 - \frac{0 - 2 \times \sum_{i=1}^{N} x^i + 2N\mu}{2\sigma^2}$$

$$0 = 2 \times \sum_{i=1}^{N} x^i - 2N\hat{\mu}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x^i}{N} = \bar{x}$$

3) Show that the maximum likelihood estimate for the distribution variance

Solution

$$\hat{\sigma^2} = argmax\left\{\frac{\partial ln(L(\mathbf{x}, \mu, \sigma^2))}{\partial \sigma^2}\right\}$$

$$\frac{\partial ln(L(\mathbf{x}, \mu, \sigma^2))}{\partial \sigma^2} = -\frac{N}{2\sigma^2} - \frac{\sum_{i=1}^{N}(x^i - \hat{\mu})^2}{2(\sigma^2)^2} \times (-1)$$

$$-\frac{N}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^{N}(x^i - \hat{\mu})^2}{2(\hat{\sigma}^2)^2} = 0$$

$$\frac{N}{\hat{\sigma}^2} = \frac{\sum_{i=1}^{N}(x^i - \hat{\mu})^2}{(\hat{\sigma}^2)^2}$$

$$N = \frac{\sum_{i=1}^{N}(x^i - \hat{\mu})^2}{\hat{\sigma}^2}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(x^i - \hat{\mu})^2}{N}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(x^i - \bar{x})^2}{N}$$

Question 7 - Programming: Implement Maximum Likelihood Estimation for a Gaussian Distribution (20 Marks)

Write a program for the following specifications (Python or Matlab preferred).
Solution:

```
Item 1 - Complete: Apply your program to the column Weight in the assignment
 dataset (see course website) and show the results.
Meam Complete: 202.270233813
Variance Complete: 238.822207499


Item 2 - GP>0: Apply your program to the column Weight conditional on GP > 0
 being true and show the results
Meam GP>0: 204.126556017
Variance GP>0: 232.349128803


Item 3 - GP<= 0: Apply your program to the column Weight conditional on GP > 0
 being false (i.e. GP = 0 ) and show the results.
Meam GP<=0: 200.85
Variance GP<=0: 239.121150794
```

Figure 2: Question 7 - Python Output

Question 8 - Programming: Implement the Naive Bayes Classifier for Hybrid Data (20 Marks+ 5 Bonus)

1) Write down the Naive Bayes classifier formula

Solution:

$$
\begin{aligned}
P(GP>0|x_1,...,x_n) &= \frac{P(x_1|GP>0)}{P(x_1|GP\leq 0)} \times .... \times \frac{P(x_n|GP>0)}{P(x_n|GP\leq 0)} \\
&= \frac{P(DraftAge|GP>0)}{P(DraftAge|GP\leq 0)} \times \frac{P(country_g|GP>0)}{P(country_g|GP\leq 0)} \times \\
&\frac{P(Height|GP>0)}{P(Height|GP\leq 0)} \times \frac{P(Weight|GP>0)}{P(Weight|GP\leq 0)} \times \\
&\frac{P(Position|GP>0)}{P(Position|GP\leq 0)} \times \frac{P(Overall|GP>0)}{P(Overall|GP\leq 0)} \times \\
&\frac{P(CSS_{rank}|GP>0)}{P(CSS_{rank}|GP\leq 0)} \times \frac{P(rs_{GP}|GP>0)}{P(rs_{GP}|GP\leq 0)} \times \\
&\frac{P(rs_G|GP>0)}{P(rs_G|GP\leq 0)} \times \frac{P(rs_A|GP>0)}{P(rs_A|GP\leq 0)} \times \frac{P(rs_P|GP>0)}{P(rs_P|GP\leq 0)} \times \\
&\frac{P(rs_{PIM}|GP>0)}{P(rs_{PIM}|GP\leq 0)} \times \frac{P(rs_{PlusMinus}|GP>0)}{P(rs_{PlusMinus}|GP\leq 0)} \times \\
&\frac{P(po_{GP}|GP>0)}{P(po_{GP}|GP\leq 0)} \times \frac{P(po_G|GP>0)}{P(po_G|GP\leq 0)} \times \\
&\frac{P(po_A|GP>0)}{P(po_A|GP\leq 0)} \times \frac{P(po_P|GP>0)}{P(po_P|GP\leq 0)} \times \\
&\frac{P(po_{PIM}|GP>0)}{P(po_{PIM}|GP\leq 0)} \times \frac{P(po_{PlusMinus}|GP>0)}{P(po_{PlusMinus}|GP\leq 0)\times}
\end{aligned}
\qquad \text{(Eq. 2)}
$$

$$
\begin{aligned}
ln(P(GP>0|x_1,...,x_n)) = {} & ln(\frac{P(DAge|GP>0)}{P(DAge|GP\leq 0)}) + ln(\frac{P(country_g|GP>0)}{P(country_g|GP\leq 0)}) + \\
& ln(\frac{P(Height|GP>0)}{P(Height|GP\leq 0)}) + ln(\frac{P(Weight|GP>0)}{P(Weight|GP\leq 0)}) + \\
& ln(\frac{P(Position|GP>0)}{P(Position|GP\leq 0)}) + ln(\frac{P(Overall|GP>0)}{P(Overall|GP\leq 0)}) + \\
& ln(\frac{P(CSS_{rank}|GP>0)}{P(CSS_{rank}|GP\leq 0)}) + ln(\frac{P(rs_{GP}|GP>0)}{P(rs_{GP}|GP\leq 0)}) + \\
& ln(\frac{P(rs_G|GP>0)}{P(rs_G|GP\leq 0)}) + ln(\frac{P(rs_A|GP>0)}{P(rs_A|GP\leq 0)}) + \\
& ln(\frac{P(rs_P|GP>0)}{P(rs_P|GP\leq 0)}) + ln(\frac{P(rs_{PIM}|GP>0)}{P(rs_{PIM}|GP\leq 0)}) + \\
& ln(\frac{P(rs_{PlusMinus}|GP>0)}{P(rs_{PlusMinus}|GP\leq 0)}) + ln(\frac{P(po_{GP}|GP>0)}{P(po_{GP}|GP\leq 0)}) + \\
& ln(\frac{P(po_G|GP>0)}{P(po_G|GP\leq 0)}) + ln(\frac{P(po_A|GP>0)}{P(po_A|GP\leq 0)}) + \\
& ln(\frac{P(po_P|GP>0)}{P(po_P|GP\leq 0)}) + ln(\frac{P(po_{PIM}|GP>0)}{P(po_{PIM}|GP\leq 0)}) + \\
& ln(\frac{P(po_{PlusMinus}|GP>0)}{P(po_{PlusMinus}|GP\leq 0))}
\end{aligned} \tag{Eq. 3}
$$

All variables are continuous, except Position and Country Group. These discrete variables were transformed into dummy variables and it was considered that all continuous variables have normal distribution.

2) Implement your Naive Bayes classifier formula to assign a class to each player. Use the draft years 1998,1999,2000 as training data and the draft year 2001 as test data. What is the test accuracy of the Naive Bayes classifier with Gaussian class-conditional distributions?

Solution:
In "assignment1_raquelaoki.py" is show each step of the Naive Bayes implementation. The train set has 711 elements and the test set 244. The Table 2 and 3 shows the confusion matrix in the test set using the Equation 2 and 3, respectively.

Table 2: Equation 2

Observed	Predicted	
	Yes	No
Yes	46 (18.85%)	59(24.18%)
No	15 (6.15%)	124 (50.82%)

Table 3: Equation 3

Observed	Predicted	
	Yes	No
Yes	31 (12.70%)	74(33.33%)
No	5 (2.44%)	134 (54.92%)

The accuracy observed in the testing set using equation 3 was 67.62% and using equation 2 was 69.67%. This difference may happen due to rounding numbers.

Bonus Question: A commonly used alternative to maximum likelihood estimation for the Gaussian variance is Bessel's bias correction. Does using s^2 instead of σ^2 change the test set accuracy, given the same training and test set?

Solution:
The implementation is also showed in "assignment1_raquelaoki.py" code. Although there are differences between the two variances, this little difference is not significant and it does not change the accuracy. So, in this case, the results are the same showed at Tables 2 and 3. An possible explanation is that the sample size is large enough to not be very influenced by the bias.