**Name: Tingyang Wei**

**Matriculation Number: G2202458H**

Serial Number: 32

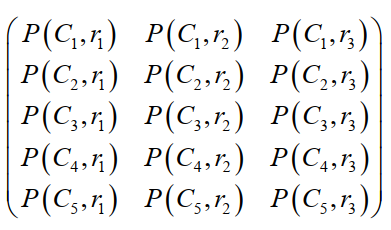
(Since my serial number is 32, the target feature herein should be PL, ***Petal Length***.)

**Problem Description:**

1. Determine the class prior probabilities, conditional probabilities and posterior probabilities for ***even bin***.
2. Subsequently verify Bayes Formula.
3. Prepare the histogram, the joint probability distribution P(C, X) as well as the P(ri|C) and P(C|i) for ***odd bins*** as a word doc.

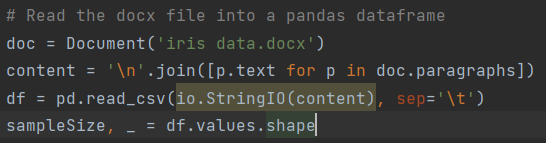
**Solution:**

To determine all the solutions related to the prior, the posterior probability, and the likelihood (conditional probabilities), a contingency table should be maintained as illustrated in the course in week 1 as **Fig.1**:



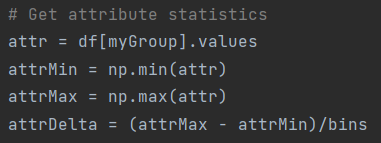
**Fig.1 Contingency Table**

Therefore, we can first construct a contingency table using Python script. Herein we apply the pandas package and docx package in Python to load the data table in the file “iris data.docx” into the DataFrame of pandas. The code can be shown as **Fig.2**:



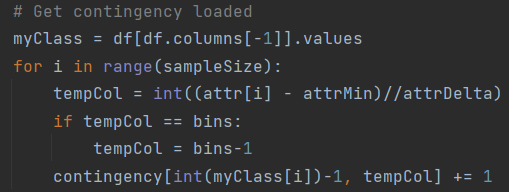
**Fig.2 Code for Reading**

Subsequently, we should determine the bin interval since it is required that the PL attribute value should be divided into 10 intervals.



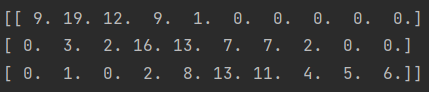
**Fig.3 Code for Dividing Attribute Value**

With the data in the DataFrame and the given attribute interval, contingency table can be constructed through bypassing all the data samples once as **Fig.4**.



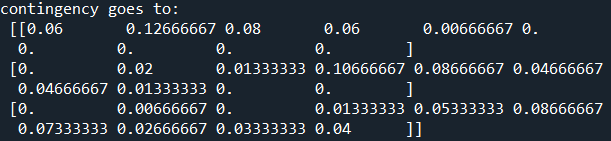
**Fig.4 Code for Constructing Contingency Table**

According to the construction, we can get a contingency table without normalization:



**Fig.5 Contingency Table with only Frequencies**

With the contingency table, other derived probability information can be naturally obtained as below:



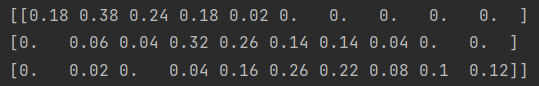
***Answer 1:***

Class prior probability :



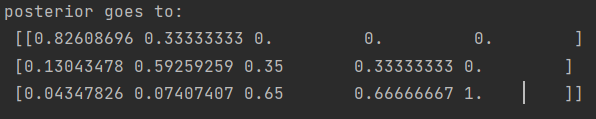
**Fig.6 Prior of Classes**

Conditional probability (Likelihood) :



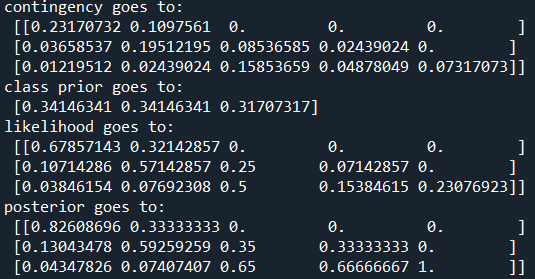
**Fig.7 Conditional Probability**

Posterior probability for ***even bins***:



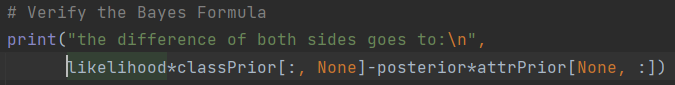
**Fig.8 Posterior Probability for even bins**

Meanwhile, I am quite confused by the problem settings here. To be specific, I don’t know why you mention the P(r|C) and conditional probability twice, considering they are basically the same thing. So, assume that you require we should sample only the even bins for all the statistical information, the results can be shown as below:

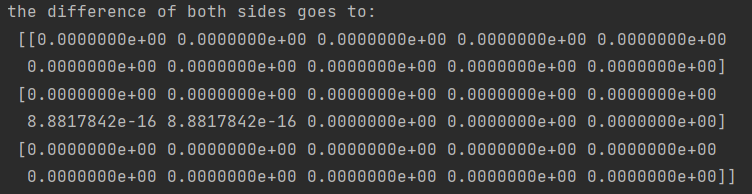


***Answer 2:***

To verify the Bayes Formula, we can verify whether the equation stands.



**Fig.9 Code for verification**

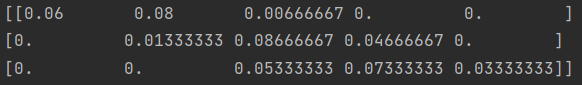


**Fig.10 The difference equals 0 and Formula gets verified**

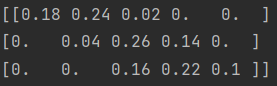
***Answer 3:***

For ***odd bins*** all the corresponding probability can be shown as below:

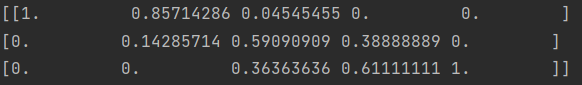
:



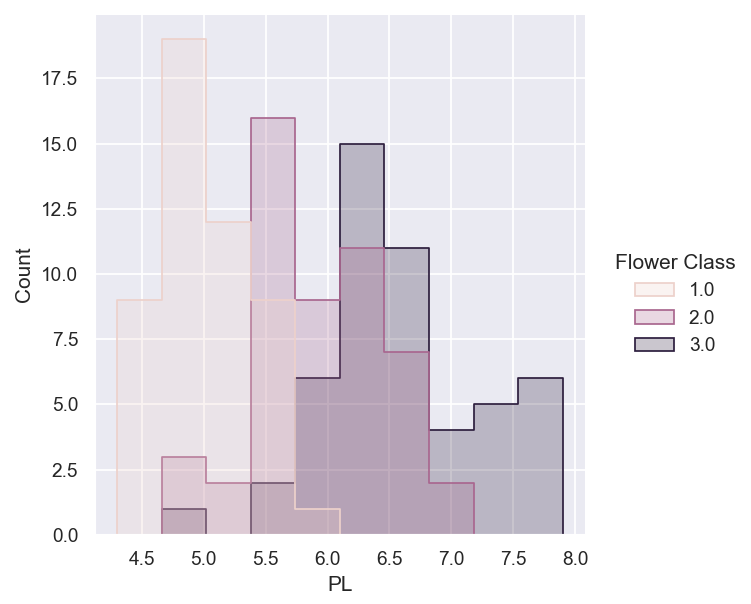
:



:



Finally, we draw the histogram on feature “PL” for the three classes using a Python package, seaborn:



Meanwhile, I am quite confused by the problem settings here. To be specific, I don’t know why you mention the P(r|C) and conditional probability twice, considering they are basically the same thing. So, assume that you require we should sample only the odd bins for all the statistical information, the results can be shown as below:

