CS259 Quantitative Methods for Computer Science
Lab 2: Bayesian Inferencing

Assessment
Our labs are assessed only formatively: you don't receive a numeric mark. Only feedback provided if submitted by the deadline. Please note however, that our quizzes may include questions based on the lab instructions, so the best way to prepare for the quizzes is to do the labs (even if only partially or submitted late).

You are not required to attend lab sessions, but it is a good opportunity to work on your assignments and get help from me and our demonstrators.

You can work on this assignment in the groups of your choice no larger than 3 students. But make sure everyone retains the copy of the code and uploads the report to myPlace, since it will be checked individually. You don't need to state the names of the students in the group.

Submission
The report for this lab assignment with your screenshots needs to be uploaded to myPlace before 10am on Tuesday October 15th as a single PDF file.

Instructions

Task 1. Bayes Theorem
You can do this task on paper and include a photo of it.
a) Applying Bayes Theorem (slides 11-13), estimate the probability of being sick based on a positive test (same task/goal as in the slides) and make sure your result is consistent with the one in the slides.
b) Same as task (a) but when the tests are 99.99% accurate.
c) Same as task (a) but when the prior chances of having the illness is 1%.

Task 2. Bayes Theorem in Java
a) Experiment with Goal 1 in our slides (Medical Diagnosis Problem) by finishing the posted template that compares the empirical probability of the patient being sick with the theoretical estimate that we obtained using Bayes Theorem during our lectures. The template checks that the relative difference between those two estimates does not exceed 10%. Please don't remove that comparison!
Hint: You may need to vary the number of trials to obtain sufficiently precise empirical estimates.
b) Same task as in (a) but when the tests are 99.99% accurate. 1(b).
Hint: You may need to change some calls to *random.nextInt()* in order to adjust your experiment to different probabilities.
Hint: You may need to update the theoretical estimate to the one you obtained in task 1(b).
c) Same task as in (a) but when the prior chances of having the illness is 1%. The theoretical estimate should come from task 1(c).

Task 3. Naïve Bayes Model

You can do this task on paper and include a photo of it.
Using Naïve Bayes Model as presented in slides 21-23, calculate the following conditional probabilities:
P(flu|cough), P(flu|fever), P(flu|sneez), P(flu|fever, sneez), P(flu|cough, fever, sneez)
Hint: you may find some helpful examples in the exercise that we did during the lecture and posted on myPlace.
The data for the model should be taken from the following table:

Cough	Fever	Sneezing	Flu
1	0	0	1
0	1	0	0
0	0	1	0
0	1	1	1
0	1	1	1
0	0	0	0

Task 4. Naïve Bayes Model in Java
a) Finish the posted template (Note: this is a different template from the one used in Task 1) so it estimates and prints the probabilities of the patient having a flu based on possible symptoms: *cough*, *fever*, and *sneezing*. You will need to insert your results from Task 3 into the appropriate places in the template. The data for the model is the same as in Task 3.

Hint: you will see in the template that rather than multiplying the feature strengths (Slides 24-25), we add the logarithms of them. This is a typical trick to avoid numeric overflow. Once all the strengths are added, we convert the sum of logarithms back to the probability.
Hint: you will also see in the template that we only use symptoms that are present, and we don't use the absence of symptoms in our model. This is to facilitate applicability of our model to MNIST later.
Hint: you will also see in the template that we apply "smoothing" to all probabilities estimated as ratios of certain counts: if those probabilities happen to be 0, we still set them to be equal to 0.01. This is to avoid dividing by zeros in our overall model.

To Submit:
Take the screenshots covering your entire code and all your outputs and add them to your report. Submit your report to myPlace as stated above under "Submission."