**CS471 Yoshii Homework 7 on Probabilities w/ PROGRAM– Based on Week 13**

**Due: Week 14 Saturday: Type your answers in black into this file. Submit this file plus Program and Test Results.**

**Total: 36 points Your score:**

**Your name: Connor Penn**

**=========================================================================**

**REVIEW QUESTIONS from the notes[10pts]:**

**Self-Test 1 Dependent Features:**

**P(cold) = priori probability of a person having a cold is 40%**

**P(fever) = priori probability of a person having a fever is 20%**

**P(cold & fever) = known to be 10%**

**P(cold | fever) = p(cold and fever) / p(fever) = 0.5**

**Self-Test 2:**

* **Let’s assume you can compute P(any word) for police arrest records after going through the training corpus. What other conditional probabilities can you compute based on these?**
* **Give specific P(word1|word2)’s you want to find out. p(evidence|crime)**
* **How will that help us?? this will help because it would improve the correlation between certain pieces of evidence and a crime. This could help police officers better detect what type of crime they are investigating based on the evidence left behind.**

**Programming Probability Collector [26pts]:**

**You must make sure your C++ program compiles with g++ and runs on empress.csusm.edu or 0pts.**

Your language alphabet is A, B, C. You want to be able to later compute P(any word).

**A program already went through the training corpus to collect trigram probabilities for:**

P(A | nothing before)

P(A | only A before)

P(A | only B before)

P(A | only C before)

P(A | AA)

P(A | AB)

P(A | AC)

P(A | BA)

…

P(B | nothing before)

P(B | only A before)

,,,,

P(C | nothing before)

P(C | only A before)

…

**Q1) How many probabilities did it have to collect? Why? [3] I had to collect 3 \* 3^0 + 3 \* 3^1 + 3 \* 3^2 = 39**

After answering this question, write a **C++ program** **to read a file** containing these probabilities and store them in some data structure. You have to **prepare your own input file** with made-up numbers. Name it **nlpProb.txt (My sample-input has the alphabet of a and b only).**

**Make sure it displays the content of the data structure in an easy to read format**. **[10]**

**Q2) How do you compute P(“abca”)? Give the equation. [3]**

**p(a| ) \* p(b | a) \* p(c|ab) \* p(a|bc) = value**

After answering this question, **extend your program** so that it will allow the user to enter any word made up of A, B, and/or C and it will compute and output its probability.

**Make sure it displays all the P’s that were used to compute the answer as I did. [10]**

(see my **demonstration program** that uses the alphabet of **a and b only**)

**Test your program** on **“abbca” “baac” and “cab”** (you should test more but these are required to be submitted) – **Must be submitted or you will not get any credit for the program!!**

**You must submit the outputs even if the programs do not work.**

**In that case, provide the bug report here: \*\***

**Submit 4 files to cougar courses:**

* **This assignment file with your answers.**
* **All sources files with good comments explaining your algorithms.**
* **nlpProb.txt (the input file)**
* **Test results (screen snapshots/recordhw script placed in one .txt file).**