IC252 Lab 1

February 21, 2019

1. **Counting.** In Himachal Pradesh, licence plates start with "HP", followed by a 2-digit RTO code, followed by N upper-case alphabets, followed by numbers between 0-9999. For example, when N=2, an example number could be "HP45 AN 4523".

   (a) Assume $N = 2$. If there are no other restrictions, how many vehicles can be registered in Himachal? Eg. "HP00 ZZ 0000" is a valid number in this case.

   (b) If the number "0" cannot be used for the RTO code and the last four digits, then how many vehicles can be registered?

   (c) In addition to the above, if "HP" cannot be used for the alphabet code, how many vehicles can be registered?

2. **Determining probabilities.** A publishing firm wants to develop special printing machines for English. For this, they need to determine the probability of occurrence of specific letters and words. You are given two large text files (fileA and fileB).

   (a) Determine the probability of each alphabet in the English language. Uppercase and lowercase alphabets are considered the same. Other characters like whitespace, punctuation and numerals are to be omitted. List the top ten occurring alphabets. Is the result same for fileA and fileB?

   (b) The measure of uncertainty is determined by its *entropy*. For example, the entropy of the arrival times for Japanese bullet trains is small (very little uncertainty, trains are almost never late.) For trains in the Indian Railways, the entropy of arrival times is larger.

   Entropy can be calculated as

   $$H = -\sum P_i \log_2 P_i$$

   where P_i is the probability of event i.

   If we consider the occurrence of alphabets in English as events of interest, determine the entropy. In other words, determine the uncertainty of alphabets in the English language. Use the results from the previous question.

   (c) How would the entropy change if all alphabets were equiprobable?

3. **Watching movies in TV.** You are at home, and the TV is available between 6 AM and 6 PM. There are several movies playing on TV and all are your favourites. Your objective is to watch the maximum number of movies. But you must watch every movie from the beginning to the end, and you cannot change the channel in between. For the movie list (with timings) below, how many movies can you watch? Also print the possible movies you can watch.
   M1 0600-0800
   M2 0800-1100
   M3 1000-1300
   M4 1500-1600
   M5 1600-1800
   Hint: The objective of this problem is to determine a subset of movies that do not overlap in time,

and this subset should have maximum number of elements.
Optional. Write a program to process any list of movies.

Input: Text files containing movies M1-MN and the times, as shown above.
Output: A subset of movies as described above, for each text file.