CS159 - In class exercise
Monday 1/31/2011

Today, we're going to be doing a few minor experiments and also trying to work through some probability problems.

To start with, get into groups of three or four. Then start reading/working through the instructions below.

I encourage you to discuss answers with other groups and to talk with them if you get stuck on any individual problem.

Please stay until at least 2:15, though I'd encourage you to stay longer if you haven't worked through all of the problems.

Administrative

- Make sure to sign the sign-up sheet before you leave.
- Don't forget that assignment 1 is due Wednesday at the beginning of class (printed out and handed to me).
- A few last minute pieces of advice for the assignment: 1) if you're having trouble coming up with the regular expressions, try and test individual components and then combine into a larger expression 2) make sure to try out (i.e. code and compile) your examples, even if I'm not asking for code and 3) I strongly suggest that you manually generate test cases and run them through each of the examples (in fact, this is a great way to start before even coding them up).
- For those who are curious about searching multiple files with grep, you can do it one of two ways. You can list multiple filenames after your regular expression, this includes things like "*" which also denote multiple files. Another option is to use the "-r" flag, which will search all files and directories starting at the specified location. For example:

  ```
  grep -E -r "your regex" .
  ```

  would search all files and subdirectories starting from the current directory.

Estimating probabilities

- Pick a team name and write it on the white board. Write the month and day of each of your birthdays under your team name. Make sure to write all of this where you will have room to write some results below it.
- Grab a deck of cards (there should be enough for one per team).
- Before we start playing with the cards, figure out in your group the answers to the following questions:
  - If you have the 13 cards from one suit (A, 2-10, J, Q, K) and let those be associated with the numbers 1-13, what is the probability of drawing a 3?
  - If we continuously draw cards and average their values, what is the expected average (for example, we saw for a dice roll it was 3.5)?
  - Let's say we now add two more aces, what is the expected average?
  - One way to calculate the expected average is:

$$E(X) = \sum_{x \in X} p(x)x$$

    where X is the set of possible values, in this case $X = 1, 2, 3, ..., 13$ and x is a particular value which has $p(x)$ probability of being chosen. Calculate the expected value using this formula. Is it the same as what you calculated before?
  - Based on this equation, does the expected value change if you now use the whole deck instead of just a single suit?
- Now, take your cards and grab 13 cards from one suit. We're going to try and estimate the probability experimentally.
  - Draw 10 random samples (with replacement) from your 13 cards and calculate the average.
  - Draw 40 more random samples (with replacement) and estimate the probability now. You can split this up among team members with another set of 13 cards in another suit.
  - Draw 50 more random samples and estimate the probability now.
  - Write your answer to these three parameter estimates (10, 50 and 100 draws) under your team name making it clear what the numbers represent.
- Now, we're going to do a similar experiment, but with the whole deck.
  - Draw 10 random samples (with replacement) from the whole deck. Count how many times you drew the ace of hearts and estimate the probability using MLE.
  - Draw 40 more random samples (with replacement) and estimate the probability now.
  - Write your answer to these under your team name and make it clear what the numbers represent.

- As the other teams finish, look at the results. Is the MLE good at estimating the probabilities? What happens as we increase the sample size? Is it better or worse for either of the experiments? Why?

Poker face

- In 5-card stud poker you're dealt 5 cards from the deck and you're supposed to make the best poker hand you can from them. The best hand you can have is a royal flush (a fairly rare event), which consists of 10-J-Q-K-A all of the same suit. If the cards are drawn from a 52 card deck, what is the probability of a royal flush being dealt?
- Let's compare this to NLP. Say we have a smallish vocabulary of 50K words and are randomly drawing words to create sentences. What is the probability that we create the five word sentence "I think therefore I am" assuming all words are equally probable and exist in the vocabulary? How does this compare to the chances of obtaining a royal flush?

Birthdays

- Looking at the birthdays on the board, is it anyone's birthday today? What is the probability that it would be someone's birthday today?
- Is anyone's birthday this week? What is the probability that this would occur?
- Are there any two people that share a birthday? What is the probability that this would happen?

Monty hall

The Monty Hall problem is based on the old game show *Let's Make a Deal*. If you are the finalist, you are presented with three doors. Behind one door is a new car; behind the other two, goats. You pick a door and then the host, who knows what is behind all of the doors, shows you a goat behind one of the doors (he always shows you a goat door). You now have the option to switch your door to the other unopened door. Should you? Specifically, what is the probability distribution of goat/car for the two doors?

- One argument is that it doesn't matter. There are now two doors left, one with a car and one with a goat, so it's 50/50. Is this a good argument?
- If you get stuck, another way to think about this problem is an extreme version of the game with 1000 doors. One door has a car and 999 have a goat. You pick a door and the host opens 998 of them with goats behind them. Should you switch? What is the probability distribution like now?

The coin game

I flip a fair coin over and over again until I get either HHH or THT. Which is more likely to come first? Why?