cs545, Spring 2012, Homework 1
Total points: 100
Due date: 5pm, Thursday, 2nd of February

Submit written or printed homeworks to our TA’s mailbox (Nisha Kiran) on the 5th floor of the computer science building.

Late policy: No late homeworks accepted.

You may discuss problems at a high level with classmates, but you must solve and write them up yourselves.

Question 1

One way of writing acronyms in English is to use a fully-capitalized three letter word, without any periods (e.g. NBA, IBM). Assume the following probabilities:

\[
P(\text{acronym} \mid \text{three-capitalized-letter-word}) = 0.9
\]
\[
P(\text{acronym} \mid \neg \text{three-capitalized-letter-word}) = 0.0001
\]
\[
P(\text{three-capitalized-letter-word}) = 0.0003
\]

We will use Bayes’ rule to compute the conditional probability that a word has three capitalized letters, given that it is an abbreviation. Show all your work.

1. Compute the probability of the event “a fully capitalized three letter word occurs and this word is an abbreviation”
2. Compute the probability of the event “an abbreviation occurs.”
3. Now, using the results from (1) and (2), compute the conditional probability that a word has three capitalized letters, given that it is an acronym.

Question 2

Suppose two coins, \( X \) and \( Y \), are flipped simultaneously, with the following joint distribution over their outcomes:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.32</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>0.48</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Determine whether the outcomes of the simultaneous flip are independent of one another, or whether the coins have undergone quantum entanglement, leading to a strange probabilistic dependency between their tosses.
Question 3

In class, we talked about a unigram language model, in which the occurrence of each word in a sentence is considered an independent event. Now, we will introduce the bigram language model. Consider a sentence of four words: \( s = w_1, w_2, w_3, w_4 \)

1. Using the chain rule, and chaining the words in the order in which they appear, write the joint probability \( P(s) \) in terms of individual conditional probabilities of the words.

2. Now, make what is called a Markov Assumption. That is, assume that each word is conditionally independent of all other words in the sentence, given the single preceding word. Rewrite your solution using this assumption.

3. Assuming that the four words in \( s \) are all unique, how many distinct probability distributions over words are being used?

Question 4

In this question we will derive the maximum likelihood estimator for the Bernoulli (coin-flip) distribution. Suppose that we possess a coin with unknown bias \( \theta \). We flip it some number of times, and observe \( H \) heads flips and \( T \) tails flips.

1. Write down the likelihood function \( L(\theta) \) (that is, the probability of our observations as a function of the unknown parameter).

2. Our goal is to compute the value for \( \theta \) which maximizes this function. We will often find it convenient to instead maximize the log-likelihood function \( LL(\theta) = \log L(\theta) \).\(^1\) By taking the log of your answer to part (1), write down the log-likelihood function.

3. By taking the derivative of this function and setting it to zero, show how to compute the maximum-likelihood \( \theta \) in terms of the observed numbers of heads and tails, \( H \) and \( T \).

\(^1\)Of course, since \( \log \) is a monotonically increasing function, the same \( \theta \) value will maximize both.