Questions for Flipped Classroom Session of COMS 4705
Week 1, Fall 2014. (Michael Collins)

Question 1

We’d like to define a language model with $V = \{\text{the, a, dog}\}$, and $p(x_1 \ldots x_n) = \gamma \times 0.5^n$ for any $x_1 \ldots x_n$, such that $x_i \in V$ for $i = 1 \ldots (n-1)$, and $x_n = \text{STOP}$, where $\gamma$ is some expression (which may be a function of $n$).

Which of the following definitions for $\gamma$ give a valid language model?

(Hint: recall that $\sum_{n=1}^{\infty} 0.5^n = 1$)

1. $\gamma = 3^{n-1}$
2. $\gamma = 3^n$
3. $\gamma = 1$
4. $\gamma = \frac{1}{3^n}$
5. $\gamma = \frac{1}{3^{n-1}}$

Question 2

In this question we consider a very simple setting, where every sentence is of length 2 (not including the STOP symbol): that is, every sentence is of the form $u,v$ where $u \in V$ and $v \in V$ for some vocabulary $V$. We define $X_1$ to be the random variable (RV) corresponding to the first word in the sentence, and $X_2$ to be the RV corresponding to the second word.

**Question:** In our first model, we assume that for any $u,v$,

$$P(X_1 = u, X_2 = v) = P(X_1 = u) \times P(X_2 = v)$$

i.e. the two random variables are independent.

For this model, prove that

$$\sum_{u \in V} \sum_{v \in V} P(X_1 = u, X_2 = v) = 1$$
**Question:** In our second model, we assume that for any $u, v$,

$$P(X_1 = u, X_2 = v) = P(X_1 = u) \times P(X_2 = v | X_1 = u)$$

i.e. the two random variables are *not independent*.

For this model, prove that

$$\sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P(X_1 = u, X_2 = v) = 1$$

**Question 3**

Nathan L. Pedant would like to build a *spelling corrector* focused on the particular problem of *there* vs *their*. The idea is to build a model that takes a sentence as input, for example

- *He saw their football in the park* (1)
- *He saw their was a football in the park* (2)

and for each instance of *their* or *there* predict whether the true spelling should be *their* or *there*. So for sentence (1) the model should predict *their*, and for sentence (2) the model should predict *there*. Note that for the second example the model would correct the spelling mistake in the sentence.

Nathan decides to use a language model for this task. Given a language model $p(w_1 \ldots w_n)$, he returns the spelling that gives the highest probability under the language model. So for example for the second sentence we would implement the rule

If $p(He \ saw \ their \ was \ a \ football \ in \ the \ park) > p(He \ saw \ there \ was \ a \ football \ in \ the \ park)$

Then Return *their*

Else Return *there*

**Question:** The first language model Nathan designs is of the form

$$p(w_1 \ldots w_n) = \prod_{i=1}^{n} q(w_i)$$
where
\[ q(w_i) = \frac{\text{Count}(w_i)}{N} \]
and Count\((w_i)\) is the number of times that word \(w_i\) is seen in the corpus, and \(N\) is the total number of words in the corpus.

Let’s assume \(N = 10,000\), Count\((\text{there})\) = 110, and Count\((\text{their})\) = 50. Assume in addition that for every word \(v\) in the vocabulary, Count\((v) > 0\). What does the rule given above return for He saw their was a football in the park? (there or their?)

Does this seem like a good solution to the there vs their problem?

**Question:** the second method that Nathan tries is to define
\[ p(w_1 \ldots w_n) = q(w_1) \prod_{i=2}^{n} q(w_i|w_{i-1}) \]
where
\[ q(w_i|w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i)}{\text{Count}(w_{i-1})} \]
and Count\((w_{i-1}, w_i)\) is the number of times that \(w_{i-1}\) is seen followed by \(w_i\) in the corpus. You can again assume that for any word \(v\) in the vocabulary, Count\((v) > 0\).

Why might this model be better than the model in the previous question?

What problems might this model have?

**Question 4**

Suppose we build a language model that makes use of a second-order Markov assumption, that is
\[ P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i|X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \]
So we assume that the \(i\)'th word \(X_i\) is independent of \(X_1 \ldots X_{i-3}\), once we condition on \(X_{i-2}\) and \(X_{i-1}\).

Give some examples in English where English grammar suggests that this independence assumption is very clearly violated.