Will Argus
004616455

LIGN 165 HW 5

Collaborated with Colman Dekker via a phone call in which we discussed the homework questions to ensure mutual understanding and gave guidance to each other. The math was mainly done individually with us just sort of checking up with each other after each problem to make sure we agreed on directions and general approach.

Problem 1. In class we gave the following equation for the bigram probability of a sequence of words $W^{(1)}, ..., W^{(k)}$:

$$Pr(W^{(1)}, ..., W^{(k)}) = \prod_{i}^{k} Pr(W^{(i)} | W^{(i-1)} = w^{(i-1)}) \tag{1}$$

Using this formula, give an expression for the bigram probability of the sentence abab, where each character is treated as a word. Try to simplify the formula as much as possible.

$$Pr(a, b, a, b) = P(a)P(b|a)P(a|b)P(b|a)$$

~~Note that:~~

$$\cancel{P(abab) \sim \text{categorical}(\vec{\theta}_{\rtimes}) \cdot \text{categorical}(\vec{\theta}_a)\, \text{categorical}(\vec{\theta}_b)\, \text{categorical}(\vec{\theta}_a)}$$

Problem 2. Let us suppose that there are two possible symbols/words in our language, a and b. There are three conditional distributions in the bigram model for this language, $Pr(W^{(i)}|W^{(i-1)} = a), Pr(W^{(i)}|W^{(i-1)} = b)$, and $Pr(W^{(i)}|W^{(i-1)} = start)$, where $start$ is the start symbol which begins any sentence. These conditional distributions are associated with the parameter vectors $\vec{\theta}_a$, $\vec{\theta}_b$, and $\vec{\theta}_{start}$, respectively (these parameter vectors were implicit in the previous problem). For the current problem, we will assume that these parameters are fixed.

Suppose that we are given a sentence $W^{(1)}, ..., W^{(k)}$. We will use the notation $n_{x \to y}$ to denote the number of times that the symbol y occurs immediately following the symbol x in the sentence. For example, $n_{a \to a}$ counts the number of times that symbol a occurs immediately following the symbol a.

Using Equation 1, give an expression for the probability of a sentence in our language:

$$Pr(W^{(1)}, ..., W^{(k)}|\vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{start}) \tag{2}$$

The expression should make use of the $n_{x \to y}$ notation defined above. (Hint: the expression should be analogous to the formula that we found for the likelihood of a corpus under a bag of words model.)

Sentence: $W^{(1)}, ..., W^{(k)}$

$$Pr(W^{(1)}, ..., W^{(k)} \mid \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{start})$$

$$= P(W^{(1)} = w^1)\, P(W^{(2)} = w^2 \mid W^{(1)} = w^1) \ldots\ldots P(W^{(k)} = w^k \mid W^{(k-1)} = w^{k-1})$$

$$\theta_a = \begin{array}{c|l} \begin{bmatrix} a \to a \\ a \to b \\ a \to start \end{bmatrix} & \begin{array}{l} \theta_a(a) \\ \theta_a(b) \\ \theta_a(start) \end{array} \end{array}$$

$$Pr(w^{1}, ..., w^{(k)} \mid \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{start}) = \theta_{start}(w^1) \cdot \theta_{w^1}(w^2) \ldots \theta_{w^{k-1}}(w^k)$$

$$\theta_b = \begin{array}{c|l} \begin{bmatrix} b \to a \\ b \to b \\ b \to start \end{bmatrix} & \begin{array}{l} \theta_b(a) \\ \theta_b(b) \\ \theta_b(start) \end{array} \end{array}$$

$$\theta_{start} = \begin{array}{c|l} \begin{bmatrix} start \to a \\ start \to b \\ start \to c \end{bmatrix} & \begin{array}{l} \theta_{start}(a) \\ \theta_{start}(b) \\ \theta_{start}(start) \end{array} \end{array}$$

$$Pr(w^{(1)}, ..., w^{(k)} \mid \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{start}) = (\theta_a(a))^{n_{a \to a}} \cdot (\theta_a(b))^{n_{a \to b}} \cdot (\theta_a(start))^{n_{a \to start}}$$
$$\cdot (\theta_b(a))^{n_{b \to a}} \cdot (\theta_b(b))^{n_{b \to b}} \cdot (\theta_b(start))^{n_{b \to start}}$$
$$\cdot (\theta_{start}(a))^{n_{start \to a}} \cdot (\theta_{start}(b))^{n_{start \to b}} \cdot (\theta_{start}(start))^{n_{start \to start}}$$

Problem 3. Let us set the parameter vectors in our bigram model as follows:

a b start (sentence ending)

$$\begin{aligned}
\text{to a} \quad \vec{\theta}_a &= (0.7, 0.2, 0.1) \\
\vec{\theta}_b &= (0.2, 0.7, 0.1) \\
\vec{\theta}_{start} &= (0.5, 0.5, 0)
\end{aligned}$$

a b start

For example, given the current symbol a, there is probability 0.7 of transitioning to the symbol a, and probability 0.2 of transitioning to the symbol b. The third term in each vector is the probability of sentence ending after that symbol. Thus, given the current symbols a or b, there is probability 0.1 of the sentence ending.

Using your answer to the previous problem and these parameter values, calculate the probability of the string $aabb$.

$n_{a\to a} = 1 \quad n_{a\to b} = 1 \quad n_{b\to a} = 0 \quad n_{b\to b} = 1$

$n_{a\to start} = 0 \quad n_{b\to start} = 1 \quad n_{start\to a} = 1 \quad n_{start\to b} = 0$

$$P(aabb) = (0.7)^1 \cdot (0.2)^1 \cdot (0.1)^0 \cdot (0.2)^0 \cdot (0.7)^1 \cdot (0.1)^1 \cdot (0.5)^1 \cdot (0.5)^0 \cdot (0)^0$$

$$= 0.0049$$

Problem 4. In the previous problem we assumed that we knew the exact values of the parameter vectors $\vec{\theta}_a$, $\vec{\theta}_b$, and $\vec{\theta}_{start}$. In the current problem, we will assume that there are actually two possible sets of parameter vectors, $\vec{\boldsymbol{\theta}_1}$ and $\vec{\boldsymbol{\theta}_2}$. We do not know ahead of time which is the correct set of parameters.

The first set of parameters $\vec{\boldsymbol{\theta}_1}$ is defined by:

$$\vec{\theta}_a = (0.7, 0.2, 0.1)$$
$$\vec{\theta}_b = (0.2, 0.7, 0.1)$$
$$\vec{\theta}_{start} = (0.5, 0.5, 0)$$

The second set of parameters $\vec{\boldsymbol{\theta}_2}$ is defined by:

$$\vec{\theta}_a = (0.2, 0.7, 0.1)$$
$$\vec{\theta}_b = (0.7, 0.2, 0.1)$$
$$\vec{\theta}_{start} = (0.5, 0.5, 0)$$

We will assume that both sets of parameters have equal prior probability: $P(\vec{\boldsymbol{\theta}_1}) = P(\vec{\boldsymbol{\theta}_2}) = 0.5$.

Compute the marginal probability of the string *aabb* given these possible sets of parameters.

$$P(aabb) = P(\theta_1)\,P(aabb \mid \theta_1) + P(\theta_2)\,P(aabb \mid \theta_2)$$

$\underline{\theta_1}$ $\quad P(\theta_1) = 0.5$

$P(aabb \mid \theta_1) = 0.0049$ (found in problem 3)

$\underline{\theta_2}$ $\quad P(\theta_2) = 0.5$

$$P(aabb/\theta_2) = (0.5)^1(0.2)^1(0.7)^1(0.2)^1(0.1)^1$$
$$= 0.0014$$

$$P(aabb) = (0.5)(0.0049) + (0.5)(0.0014) = 0.00315$$

Problem 5. In the current problem, we will try to address a learning problem: determining which of the parameters $\vec{\theta_1}$ or $\vec{\theta_2}$ is the correct one. Using the marginal probability that you computed in Problem 4, compute the posterior probability $P(\vec{\theta_1}|aabb)$.

The quantity that you computed should be greater than $P(\vec{\theta_2}|aabb)$. Why is this true?

$$P(\theta_1 \mid aabb) = \frac{P(aabb \mid \theta_1)\, P(\theta_1)}{P(aabb)}$$

$$= \frac{0.0049 \cdot 0.5}{0.00315} = 0.7\overline{7}$$

$$P(\theta_2 \mid aabb) = \frac{P(aabb \mid \theta_2)\, P(\theta_2)}{P(aabb)}$$

$$= \frac{(0.0014)(.5)}{0.00315} = 0.2\overline{2}$$

$P(\theta_1 | aabb)$ is greater than $P(\theta_2 | aabb)$ since θ_1 and θ_2 are equally likely and θ_1 is more likely to produce sentence aabb, hence $P(\theta_1 | aabb)$ is greater

Problem 6. Find a string c (consisting of a's and b's) such that $P(\vec{\boldsymbol{\theta}}_\mathbf{1}|c)$ is greater than the value $P(\vec{\boldsymbol{\theta}}_\mathbf{1}|aabb)$ that you found in the previous problem. How did you construct this string?

string c = aaaa

$$P(\theta_1 | c) = \frac{P(aa | \theta_1)\, P(\theta_1)}{P(aa)}$$

$$P(aa|\theta_1) = (0.5)^1 (0.7)^1 (0.7)(0.7)(0.1) = 0.01715$$

$$P(aa|\theta_2) = (0.5)(0.2)(0.2)(0.2)(0.1) = 0.0004$$

$$P(aa) = 0.008775$$

$$P(\theta_1|c) = \frac{(0.01715)(.5)}{0.008775} = 0.977$$

$$P(\theta_2|c) = \frac{(0.0004)(.5)}{0.008775} = 0.0228$$

$$\underline{P(\theta_1 | aaaa) > P(\theta_1 | aabb)}$$

Since for θ_1, a followed by a is the most likely, a string of all a's would make θ_1 more likely than θ_2 by a large margin, shown above. So 4 a's in a row made $P(\theta_1 | aaaa) > P(\theta_1 | aabb)$

Problem 7. Find a string c such that $P(\vec{\theta_1}|c) < P(\vec{\theta_2}|c)$. How did you construct this string?

$c = ab$

$$P(c|\theta_2) = (0.5)(0.7)(0.1) = 0.35$$

$$P(c|\theta_1) = (.5)(.2)(.1) = 0.01$$

$$P(c) = (0.5)(0.35) + (0.5)(0.01) = 0.18$$

$$P(\theta_1|c) = \frac{P(c|\theta_1)P(\theta_1)}{P(c)} = \frac{(0.01)(0.5)}{0.18} = 0.028$$

$$P(\theta_2|c) = \frac{P(c|\theta_2)P(\theta_2)}{P(c)} = \frac{(0.35)(0.5)}{0.18} = 0.9722$$

Clearly $P(\theta_2|c) > P(\theta_1|c)$.

$c = ab$ was chosen as the transition from a to b is likely for θ_2 but unlikely for θ_1, hence for string c, θ_2 would be more likely than θ_1, as is the case shown above.