LING/COMP 445, LING 645
Problem Set 5

Name: Luna Dana , **McGill ID**: 260857641

Due before 10:05 AM on Tuesday, November 23, 2021

Please enter your name and McGill ID above.

This problem set consists only of questions involving mathematics or English or or a combination of the two (no coding questions this time). Please put your answers in an answer box like in the example below.

Once you have entered your answers, please compile your copy of this LaTeX into a PDF and submit

(i) the compiled PDF renamed to `ps5-lastname-firstname.pdf` and
(ii) the raw LaTeX file renamed to `ps5-lastname-firstname.tex`

to the Problem Set 5 folder under 'Assignments' on MyCourses.

Example Problem: This is an example question using some fake math like this $L = \sum_0^\infty \mathcal{G}\delta_x$.

Example Answer: Put your answer in the box provided, like this:

Example answer is $L = \sum_0^\infty \mathcal{G}\delta_x$.

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

$$\Pr(W^{(1)}, \ldots, W^{(k)}) \stackrel{\text{def}}{=} \prod_{i}^{k} \Pr(W^{(i)}|W^{(i-1)} = w^{(i-1)}) \tag{1}$$

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.

Important note: Throughout this problem set, the vocabulary will be $V \stackrel{\text{def}}{=} \{a, b\}$. We will assume the length of the sentence is fixed at some k, and *we will not use the stop symbol.* That is, in a sentence of length k, for $1 \le i \le k$, the possible values for the random variable $W^{(i)}$ are just a and b, and we will refer to the beginning of the string as $W^{(0)} = \rtimes$ always. So, $\Pr(W^{(1)} = a \mid W^{(0)} = \rtimes)$ is the probability that the string starts with a.

Answer 1: Please put your answer in the box below.

Since we are looking for the probability of the sentence *abab* to appear in this order we know that : $w^{(0)} = \rtimes$ (representing the beggining of a string) , $w^{(1)} = a, w^{(2)} = b, w^{(3)} = a$, and $w^{(4)} = b$.

Thus, $\Pr(\rtimes abab \ltimes)$

$= \Pr(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = \Pr(W^{(1)} = w^{(1)}, W^{(2)} = w^{(2)}, W^{(3)} = w^{(3)}, W^{(4)} = w^{(4)})$

$= \Pr(W^{(1)} = a, W^{(2)} = b, W^{(3)} = a, W^{(4)} = b)$

$= \prod_i^4 \Pr(W^{(i)}|W^{(i-1)} = w^{(i-1)})$

$= \Pr(W^{(1)} = a|W^{(0)} = \rtimes) \cdot \Pr(W^{(2)} = b|W^{(1)} = a) \cdot \Pr(W^{(3)} = a|W^{(2)} = b) \cdot \Pr(W^{(4)} = b|W^{(3)} = a)$

$= \Pr(a|W^{(0)} = \rtimes) \cdot \Pr(b|W^{(1)} = a) \cdot \Pr(a|W^{(2)} = b) \cdot \Pr(b|W^{(3)} = a)$

$= \Pr(a|W^{(0)} = \rtimes) \cdot \Pr(b|W^{(1)} = a)^2 \cdot \Pr(a|W^{(3)} = b)$

Problem 2: 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 $\Pr(W^{(i)}|W^{(i-1)} = \rtimes)$. These conditional distributions are associated with the parameter vectors $\vec{\theta}_a$, $\vec{\theta}_b$, and $\vec{\theta}_\rtimes$, respectively (these parameter vectors were implicit in the previous problem). For the current problem, we will assume that these parameters are fixed. Use a second subscript notation to denote components of these vectors, so $\theta_{ab} = \Pr(W^{(i)} = b \mid W^{(i-1)} = a)$.

Suppose that we are given a sentence $W^{(1)}, \ldots, W^{(k)}$. 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 length k sentence in our language:

$$\Pr(W^{(1)}, \ldots, W^{(k)}|\vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_\rtimes)$$

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.

Answer 2: Please put your answer in the box below.

We know that if n_w is the count of the word w from the vocabulary in the sentence, then the probability of the sentence can be written as

$$\Pr(w^{(1)}, \ldots, w^{(k)}) = \prod_{w \in V} \theta_w^{n_w} \tag{2}$$

Moreover, let's define V_a, V_b and $V_{\rtimes}$ which are subsets of the vocabulary $V = \{a, b\}$ where for $i \in \{a, b, \rtimes\}$, i is the preceding word of V_i.

$$\Pr(W^{(1)}, \ldots, W^{(k)} | \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}) = \Pr(V_a | \vec{\theta}_a) \times \Pr(V_b | \vec{\theta}_b) \times \Pr(V_{\rtimes} | \vec{\theta}_{\rtimes}) \tag{3}$$

Hence,

$$\Pr(W^{(1)}, \ldots, W^{(k)} | \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}) = \prod_{w \in V_a} \theta_{aw}^{n_{a \to w}} \times \prod_{w \in V_b} \theta_{bw}^{n_{b \to w}} \times \prod_{w \in V_{\rtimes}} \theta_{\rtimes w}^{n_{\rtimes \to w}} \tag{4}$$

$$\Pr(W^{(1)}, \ldots, W^{(k)} | \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}) = \prod_{i \in \{a, b, \rtimes\}} \prod_{w \in V_i} \theta_{iw}^{n_{i \to w}} \tag{5}$$

Problem 3: Assume the parameter vectors in our bigram model have the following values:

$$\vec{\theta}_a = (0.7, 0.3)$$
$$\vec{\theta}_b = (0.2, 0.8)$$
$$\vec{\theta}_{\rtimes} = (0.5, 0.5)$$

The first vector indicates that if the current symbol a, there is probability 0.7 of transitioning to the symbol a, and probability 0.3 of transitioning to the symbol b. Using your answer to the previous problem and these parameter values, calculate the probability of the string $aabab$.

Answer 3: Please put your answer in the box below.

$\Pr(\rtimes aabab \ltimes) = \theta_{\rtimes a} \theta_{aa} \theta_{ab} \theta_{ba} \theta_{ab} = (0.5)(0.7)(0.3)(0.2)(0.3) = 0.0063$

Problem 4: In Problem 2, you found an expression for the bigram probability of a sentence in our language, which contains the symbols a and b. In that problem, we assumed that there were fixed parameter vectors $\vec{\theta}$ associated with each conditional distribution. In this problem, we will consider the setting in which we have uncertainty about the value of these parameters.

As we did in class, we will use the Dirichlet distribution to define a prior over parameters. Assume each parameter vector is drawn independently given $\vec{\alpha}$:

$$\vec{\theta}_{\mathbf{c}} \mid \vec{\alpha} \sim \text{Dirichlet}(\vec{\alpha}) \tag{6}$$
$$w^{(i)} \mid w^{(i-1)} \sim \text{categorical}(\vec{\theta}_{w^{(i-1)}}) \tag{7}$$
$$w^{(1)} \sim \text{categorical}(\vec{\theta}_{\rtimes}) \tag{8}$$

Suppose that we have a fixed-length sentence $S = W^{(1)}, \ldots, W^{(k)}$. Give an expression for the joint probability $\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes} | \vec{\alpha})$ using the definitions of Dirichlet distributions and likelihoods we defined in class.

Answer 4: Please put your answer in the box below.

The Dirichlet distribution can be defined as follows (from Chapter 16 Hierarchical Models) :

$$\Pr\left(\vec{\theta}|\vec{\alpha}\right) = \frac{1}{B(\vec{\alpha})}\prod_{i=1}^{K}\theta_i^{\alpha_i-1} = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}\theta_i^{\alpha_i-1}$$

Recall V_a, V_b and $V_{\rtimes}$ which are subsets of the vocabulary $V = \{a, b\}$ where for $i \in \{a, b, \rtimes\}$, i is the preceding word of V_i.

Therefore for $i \in \{a, b, \rtimes\}$,

$$\Pr\left(\vec{\theta}_i|\vec{\alpha}\right) = \frac{1}{B(\vec{\alpha})}\prod_{w\in V_i}\theta_{i\to w}^{\alpha_w-1} = \frac{\Gamma(\sum_{w\in V_i}\alpha_w)}{\prod_{w\in V_i}\Gamma(\alpha_w)}\prod_{w\in V_i}\theta_{i\to w}^{\alpha_w-1}$$

Using the chain rule :

$$\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha}) = \Pr(S, \vec{\theta}_a, |\vec{\alpha}) \times \Pr(S, \vec{\theta}_b, |\vec{\alpha}) \times \Pr(S, \vec{\theta}_{\rtimes}|\vec{\alpha})$$

$$\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha}) = \Pr(S|\vec{\theta}_a) \times \Pr(\vec{\theta}_a|\vec{\alpha}) \times \Pr(S|\vec{\theta}_b) \times \Pr(\vec{\theta}_b|\vec{\alpha}) \times \Pr(S|\vec{\theta}_{\rtimes}) \times \Pr(\vec{\theta}_{\rtimes}|\vec{\alpha})$$

Simplifying we get,

$$\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha}) = \Pr(S|\vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}) \times \Pr(\vec{\theta}_a|\vec{\alpha}) \times \Pr(\vec{\theta}_b|\vec{\alpha}) \times \Pr(\vec{\theta}_{\rtimes}|\vec{\alpha})$$

$$\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha}) = \frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)} \times (\prod_{w\in V_a}\theta_{a\to w}^{[n_{a\to w}+\alpha_w]-1}\prod_{w\in V_b}\theta_{b\to w}^{[n_{b\to w}+\alpha_w]-1}\prod_{w\in V_{\rtimes}}\theta_{\rtimes\to w}^{[n_{\rtimes\to w}+\alpha_w]-1}) \quad (9)$$

Problem 5: In the previous problem, you found a formula for the joint probability of a sentence and a set of bigram model parameters. Using this, give a formula for the marginal probability of the sentence $\Pr(S|\vec{\alpha})$.

Hint: The formula should be analogous to the formula derived in class for marginal probability of a corpus under a bag of words model. Whereas before there was only a single parameter vector $\vec{\theta}$, now there are three parameter vectors that need to be marginalized away. Otherwise the calculation will be similar.

Answer 5: Please put your answer in the box below.

In class, we saw that :

$$\Pr(\mathbf{C}|\vec{\alpha}) = \int_{\Theta}\Pr(C, \vec{\theta})$$

We also know that the following formula holds :

$$\int_{\Theta}\prod_{i=1}^{K}\theta_i^{x_i-1} = \frac{\prod_{i=1}^{K}\Gamma(x_i)}{\Gamma(\sum_{i=1}^{K}x_i)} \quad (10)$$

Now, let's give a formula for the marginal probability of the sentence

$$\Pr(\mathbf{S}|\vec{\alpha}) = \int_{\Theta_a}\int_{\Theta_b}\int_{\Theta_{\rtimes}} \Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha}) d\vec{\theta}_{\rtimes} d\vec{\theta}_b d\vec{\theta}_a$$

We first replace $\Pr(S, \vec{\theta}_a, \vec{\theta}_b, \vec{\theta}_{\rtimes}|\vec{\alpha})$ with what we found in question 4. We can factor $\frac{\Gamma(\sum_{w\in V}\alpha_w)}{\prod_{w\in V}\Gamma(\alpha_w)}$ since it is not depending on $\Theta_{\rtimes}$, Θ_a nor Θ_b.

$$\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)} \int_{\Theta_a}\int_{\Theta_b}\int_{\Theta_{\rtimes}} \Big(\prod_{w\in V_a} \theta_{a\to w}^{[n_{a\to w}+\alpha_w]-1} \prod_{w\in V_b} \theta_{b\to w}^{[n_{b\to w}+\alpha_w]-1} \prod_{w\in V_{\rtimes}} \theta_{\rtimes\to w}^{[n_{\rtimes\to w}+\alpha_w]-1}\Big)$$

Since the 3 product respectively depend on a, b and $\rtimes$. The other 2 variables will be treated as constants during the integration and we can therefore split the triple integral as such :

$$\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)} \int_{\Theta_a} \Big(\prod_{w\in V_a} \theta_{a\to w}^{[n_{a\to w}+\alpha_w]-1}\Big) \times \int_{\Theta_b} \Big(\prod_{w\in V_b} \theta_{b\to w}^{[n_{b\to w}+\alpha_w]-1}\Big) \times \int_{\Theta_{\rtimes}} \Big(\prod_{w\in V_{\rtimes}} \theta_{\rtimes\to w}^{[n_{\rtimes\to w}+\alpha_w]-1}\Big)$$

Computing each product with gives us :

$$\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)} \frac{\prod_{w\in V_{\rtimes}}\Gamma([n_{\rtimes\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_{\rtimes}}[n_{\rtimes\to w}+\alpha_w]\right)} \frac{\prod_{w\in V_b}\Gamma([n_{b\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_b}[n_{b\to w}+\alpha_w]\right)} \frac{\prod_{w\in V_a}\Gamma([n_{a\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_a}[n_{a\to w}+\alpha_w]\right)}$$

$$\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)} \prod_{i\in a,b\rtimes} \frac{\prod_{w\in V_i}\Gamma([n_{i\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_i}[n_{i\to w}+\alpha_w]\right)}$$

Problem 6: Let us assume that the parameters of the Dirichlet distribution are $\vec{\alpha} = (1, 1)$. Using your solution to the previous problem, write an expression for $\Pr(S = aabab \mid \vec{\alpha} = (1, 1))$, the marginal probability of the string *aabab*. The expression should should contain the gamma function $\Gamma(\cdot)$. Using the properties of the gamma function discussed in class (i.e., it's relationship to the factorial) or an online calculator, compute a numerical value for this expression.

Answer 6: Please put your answer in the box below.

We have to compute :

$$\Pr(\mathbf{S}|\vec{\alpha}) = \Pr(\rtimes aabab|(1, 1))$$

We have the following results:

- $n_{a\to a} = 1$, $n_{a\to b} = 2$, $n_{b\to b} = 0$, $n_{b\to a} = 1$, $n_{\rtimes\to a} = 1$, $n_{\rtimes\to b} = 0$ where for every x and y, $n_{x\to y}$ denote the number of times that y occurs immediately followin x in the sentence.
- $V_a = \{a, b\}$, $V_b = \{a\}$, and $V_{\rtimes} = \{a\}$ which are subsets of the vocabulary $V = \{a, b\}$ where for $i \in \{a, b, \rtimes\}$, i is the preceding word of V_i.
- The Gamma function is defined as follow : $\Gamma(n) = (n-1)! \quad \forall n \in N$

We now plug the results into the following formula :

$$\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)}\frac{\prod_{w\in V_{\rtimes}}\Gamma([n_{\rtimes\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_{\rtimes}}[n_{\rtimes\to w}+\alpha_w]\right)}\frac{\prod_{w\in V_b}\Gamma([n_{b\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_b}[n_{b\to w}+\alpha_w]\right)}\frac{\prod_{w\in V_a}\Gamma([n_{a\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_a}[n_{a\to w}+\alpha_w]\right)}$$

$$\frac{\Gamma(5)}{\Gamma(1)^2}\frac{\Gamma(2)}{\Gamma(2)}\frac{\Gamma(2)}{\Gamma(2)}\frac{\Gamma(2)\Gamma(3)}{\Gamma(5)}=\frac{\Gamma(5)}{\Gamma(1)^2}\frac{\Gamma(2)\Gamma(3)}{\Gamma(5)}=\frac{1}{\Gamma(1)^2}\Gamma(2)\Gamma(3)=\frac{1}{1}(1\times 2)=2$$

Problem 7: Suppose that we have observed a sentence $S=W^{(1)},\dots,W^{(k)}$. Find an expression for the posterior distribution over the model parameters, $\Pr(\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}\mid S,\vec{\alpha})$.

Hint: Use the joint probability that you computed in Problem 4 and Bayes' rule. The solution should be analogous to the posterior probability for the bag of words model.

Answer 7: Please put your answer in the box below.

Here is the definition of Baye's rule as seen in class :

$$\Pr(\vec{\theta}|S,\vec{\alpha})=\frac{\Pr(S,\vec{\theta}|\vec{\alpha})}{\Pr(S|\vec{\alpha})}=\frac{\Gamma\left(\sum_{w\in V}[n_w+\alpha_w]\right)}{\prod_{w\in V}\Gamma([n_w+\alpha_w])}\prod_{w\in V}\theta_w^{[n_w+\alpha_w]-1}$$

Therefore using our answers to question 4 and 5 and applying Baye's rule defined just above we get :

$$\Pr(\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}\mid S,\vec{\alpha})=\frac{\Pr(S,\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}|\vec{\alpha})}{\Pr(S|\vec{\alpha})}$$

$$=\frac{\frac{\Gamma(\sum_{w\in V}\alpha_w)}{\prod_{w\in V}\Gamma(\alpha_w)}\times(\prod_{w\in V_a}\theta_{a\to w}^{[n_{a\to w}+\alpha_w]-1}\prod_{w\in V_b}\theta_{b\to w}^{[n_{b\to w}+\alpha_w]-1}\prod_{w\in V_{\rtimes}}\theta_{\rtimes\to w}^{[n_{\rtimes\to w}+\alpha_w]-1})}{\frac{\Gamma(\sum_{w\in V}\alpha_w)}{\prod_{w\in V}\Gamma(\alpha_w)}\prod_{i\in a,b\rtimes}\frac{\prod_{w\in V_i}\Gamma([n_{i\to w}+\alpha_w])}{\Gamma\left(\sum_{w\in V_i}[n_{i\to w}+\alpha_w]\right)}}$$

Problem 8: Using your formula for the posterior distribution and setting $\vec{\alpha}=(1,1)$, given an expression for the posterior distribution over parameters given the sentence *aabab*. There should be an easy way to interpret the posterior distribution, and how it was derived from the observed sentence. What is this interpretation?

Answer 8: Please put your answer in the box below.

Let's compute :

$$\Pr(S|\vec{\alpha})=\Pr(\rtimes aabab|(1,1))=\Pr(S,\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}|\vec{\alpha})=\Pr(\rtimes aabab,\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}|(1,1))$$

From question 4,

$$\Pr(S,\vec{\theta_a},\vec{\theta_b},\vec{\theta}_{\rtimes}|\vec{\alpha})=\frac{\Gamma\left(\sum_{w\in V}\alpha_w\right)}{\prod_{w\in V}\Gamma(\alpha_w)}\times(\prod_{w\in V_a}\theta_{a\to w}^{[n_{a\to w}+\alpha_w]-1}\prod_{w\in V_b}\theta_{b\to w}^{[n_{b\to w}+\alpha_w]-1}\prod_{w\in V_{\rtimes}}\theta_{\rtimes\to w}^{[n_{\rtimes\to w}+\alpha_w]-1})$$

$$=\frac{\Gamma(1)^2}{\Gamma(5)}\times\theta_{a\to a}\times\theta_{a\to b}^2\times\theta_{b\to a}\times\theta_{\rtimes\to a}=\frac{1}{12}\times\theta_{a\to a}\times\theta_{a\to b}^2\times\theta_{b\to a}\times\theta_{\rtimes\to a}$$

The idea behind the Bayesian posterior predictive distribution is that we might not want to commit to a single hypothesis in our posterior distribution, but instead might want to average over all of our posterior values.

We can therefore average all our posterior values which yields to a better prediction of the next step in the case of a model change.

Problem 9: Consider the language $L = \{a^*ba^*\}$, that is, the language consisting of some number of a's, followed by a single b, followed by some number of a's. Show that this language is not strictly 2-local.

Hint: use n-Local Suffix Substitution Closure (n-SSC).

Answer 9: Please put your answer in the box below.

A language L is a strictly n-local language if and only if for every $u_1, u_2, v_1, v_2 \in V^*$ and for every $x \in V^{n-1}$ if u_1xv_1 is in L and u_2xv_2 is in L then so is u_1xv_2 by the use n-Local Suffix Substitution Closure (n-SSC) rule seen in class.

Consider the following counter example :

- Choose a pivot of length $2 - 1 = 1$ such as $x = a$.
- Let $u_1 = aab$ and $v_1 = aa$.
- Let $u_2 = a$ and $v_2 = ba$.

We see that $u_1xv_1 = aabaaa$ and $u_2xv_2 = aaba$ are in L. However, $u_1xv_2 = aababa$ is not in L are we can therefore conclude that $L = \{a^*ba^*\}$ is not strictly 2-local.

Problem 10: Consider the language $L = \{a^n b^m c^n d^m \; |n, m \in \mathbb{N}\}$, that is, the language consisting of n a's followed by m b's followed by n c's followed by m d's where n and m are natural numbers. Show that this language is not strictly 2-local.

Hint: use the same property as in the problem above.

Answer 10: Please put your answer in the box below.

Recall that a language L is a strictly n-local language if and only if for every $u_1, u_2, v_1, v_2 \in V^*$ and for every $x \in V^{n-1}$ if u_1xv_1 is in L and u_2xv_2 is in L then so is u_1xv_2 by the use n-Local Suffix Substitution Closure (n-SSC) rule seen in class.

Consider the following counter example :

- Choose a pivot of length $2 - 1 = 1$ such as $x = b$.
- Let $u_1 = aa$ and $v_1 = ccd$.
- Let $u_2 = aaab$ and $v_2 = cccdd$.

We see that $u_1xv_1 = aabccd$ and $u_2xv_2 = aaabbcccdd$ are in L. However, $u_1xv_2 = aabcccdd$ is not in L are we can therefore conclude that $L = \{a^n b^m c^n d^m \; |n, m \in \mathbb{N}\}$ is not strictly 2-local.

Problem 11: Show that the language $L = \{a^n b^m c^n d^m \; |n, m \in \mathbb{N}\}$ is not strictly k-local, for any value of k.

Answer 11: Please put your answer in the box below.

Recall that a language L is a strictly n-local language if and only if for every $u_1, u_2, v_1, v_2 \in V^*$ and for every $x \in V^{n-1}$ if u_1xv_1 is in L and u_2xv_2 is in L then so is u_1xv_2 by the use n-Local Suffix Substitution Closure (n-SSC) rule seen in class.

Consider the following counter example :

- Choose a pivot of length $k-1$ such as $x = b^{k-1}$.
- Let $u_1 = a^n$ and $v_1 = c^n d^{k-1}$.
- Let $u_2 = a^n b$ and $v_2 = c^n d^k$.

We see that $u_1xv_1 = a^n b^{k-1} c^n d^{k-1}$ and $u_2xv_2 = a^n b b^{k-1} c^n d^k = a^n b^k c^n d^k$ are in L. However, $u_1xv_2 = a^n b^{k-1} c^n d^k$ is not in L are we can therefore conclude that $L = \{a^n b^m c^n d^m \; | n, m \in \mathbb{N}\}$ is not strictly k-local.

Problem 12: In class we proved that $k-\text{SSC}(L) \implies L \in \text{SL}_k$. In other words, if a language satisfies k-Local Suffix Substitution Closure, then it is k-strictly local.

Use this theorem to prove that k-strictly local languages are closed under intersection. More precisely, prove that if $L_1 \in \text{SL}_k$ and $L_2 \in \text{SL}_k$, then $L_1 \cap L_2 \in \text{SL}_k$.

Answer 12: Please put your answer in the box below.

$L_1 \in \text{SL}_k$ and $L_2 \in \text{SL}_k$ are k-strictly local which means that for every $u_1, u_2, v_1, v_2 \in V^*$ and for every $x \in V^{n-1}$ if u_1xv_1 is in L_1 (same applies for L_2) and u_2xv_2 is in L_1 (same applies for L_2) then so is u_1xv_2 by the use n-Local Suffix Substitution Closure (n-SSC) rule seen in class.

Since $L_1 \in \text{SL}_k$, we know that if $u_1xv_1 \in L_1 \cap L_2$ and $u_2xv_2 \in L_1 \cap L_2$ then $u_1xv_2 \in L_1 \cap L_2$ because $L_1 \cap L_2 \in L_1$.

Since $L_2 \in \text{SL}_k$, we know that if $u_1xv_1 \in L_1 \cap L_2$ and $u_2xv_2 \in L_1 \cap L_2$ then $u_1xv_2 \in L_1 \cap L_2$ because $L_1 \cap L_2 \in L_2$.

Therefore, k-strictly local languages are closed under intersection and our statement is prooved.