Questions

The provided code is set up to use a uniform frequency distribution (unifreqs) which assumes each verb type has the same chance of being observed (line 147). This is not what actually happens during language learning since verbs are observed with different rates. Modify the code to use the pastfreqs distribution instead (this is an estimate of how often each past tense is observed) and rerun the code to answer the following questions.

PART 1

Q1

What nonce prediction rate for the regulars does the model predict after 50,000 iterations if observations are instead sampled from the past tense frequency distribution?

Past tense frequency: 81%

Q2

Albright and Hayes (2003) found that adults produced the regular past tense 81.5% of the time in their wug test. Which distribution (uniform or past) better matches these results? Why might this be? (hint: consider the proportion of examples that are regular in each distribution)

Past distribution better matches these results. This is because past frequency distribution reflects what input adults actually recieve, whereas uniform distribution assigns the same frequency of one to all types of verbs.

PART 2

Q1

Regular verbs reach high accuracy very quickly, even infrequent verbs like ‘crimp/crimped’ are already 80% accurate by about 2000 iterations. Explain why ‘crimp/crimped’ is so accurate so quickly. Specifically, discuss the roles of the general constraints and scale for achieving correct predictions on this verb.

‘crimp/crimped’ becomes accurate so quickly because it is part of the REG class of verbs. Verbs belonging to the REG class occur more frequently in the training data, so the weights for the REG constraint will update more quickly than for other general constraints. Therefore, once the model does encounter an infrequent regular verb like ‘crimp/crimped’, it already is biased towards verbs of the REG class that it belongs to. At this point, the scale for that particular verb merely fine-tunes its accuracy, as the general constraints have already allowed it to reach high accuracy very quickly.

Q2

For irregular verbs, the scales will need to reach a strength of about 5 for these verbs to be produced at around 80% accuracy. Examine how accuracy changes for the most frequent verb TINk->T$t over time. At (roughly) what iteration does the model predict 80% accuracy for TINk->T$t? Explain why this irregular verb takes so much longer to reach 80% accuracy than ‘crimp/crimped’. And also, discuss the roles of the general constraints and scale for achieving correct predictions on this verb.

At roughly 45000 iterations, TINk->T$t reaches 80% accuracy.
This is because unlike ‘crimp/crimped’, TINk->T$t does not belong to the REG class. Again, verbs belonging to the REG class are more frequent, so the REG constraint updates quicker than for other constraints, including the X->$t constraint that applies to TINk->T$t. Since the general constraints cause the model to favor the verb following the regular past tense rule, irregular verbs such as TINk->T$t must rely on their verb-specific scales to reach high accuracy. Even though TINk->T$t occurs frequently, it takes many iterations for it to reach 80% accuracy through updates to its scale.

Q3
Although ‘crimp/crimped’ gets to about 80% accuracy very quickly, it then takes a long time for it to improve beyond this and get to 90% accuracy and above. Which verb, ‘crimp/crimped’ or ‘think/thought’, gets to above 90% first? Why is this verb mastered earlier?
‘think/thought’ gets to above 90% accuracy first.
Although ‘crimp/crimped’ initially reaches 80% accuracy much faster than ‘think/thought’ due to the general constraints, it takes longer to get to 90% due to its infrequency. Since ‘crimp/crimped’ occurs less frequently, it will be sampled less often, meaning its scale will update less often as well. Thus, it takes many more iterations for it to improve beyond 80% and surpass 90% accuracy.
On the other hand, although ‘think/thought’ takes longer to reach 80% due to the general constraints strongly preferring the REG class, it reaches 90% faster due to its frequency. It is the most frequent verb, so it is sampled more often, meaning its scale will update more frequently as well, allowing it to be mastered earlier than an infrequent regular verb like ‘crimp/crimped’.

PART 3
Now that you understand the basic mechanics of the model, in this part of the assignment, we will examine whether the model captures some observations about children’s acquisition of the English past tense.

Questions
Each of the questions below asks you to analyze the model to determine whether it captures a different observation about acquisition.

Type Frequency
All else being equal, children tend to acquire verbs belonging to larger classes (classes with more types) earlier than verbs belonging to smaller classes. To test the model’s predictions for this, we will consider verbs from three differently-sized classes: go/went (with 3 verbs in its class), set/set (with 35 verb types in the NULL class), and use/used (with 4041 verb types in the REG class). To make sure the token frequency of these verbs isn’t affecting the predictions, we will artificially set the frequencies of these three verbs to the same value: 3099 (the frequency of ‘go/went’). To do this, after setting the frequency dictionary to the past tense frequencies, set the frequency of use/used and set/set to 3099. Then run the model several times to examine how quickly it learns each

of the three verbs and answer the following questions. Undo these modifications once you're done with this part.

Q1

At (roughly) what iteration does the model reach 80% accuracy for each of the three verbs: go/went, set/set, and use/used? Explain whether this order corresponds to these classes' type frequency.

go/went: ~61900 iterations
set/set: ~69900 iterations
use/used: ~1200 iterations
use/used reaches 80% accuracy much faster than go/went and set/set. This is because use/used is part of the REG class, which has a much higher type frequency than the g5 -> wEnt and NULL classes, which go/went and set/set belong to, respectively. A specific general constraint will be updated more frequently if more verbs belong to that class– meaning the higher the type frequency, the faster verbs of that class will reach 80% accuracy. The REG class has the highest type frequency, so verbs that belong to that class, such as use/used, will reach 80% the fastest. Smaller classes such as the NULL and g5 -> wEnt classes, must rely on scales to grow their accuracies, which take many iterations to reach 80% accuracy. The NULL class, which set/set belongs to, has a type frequency of 35, so it reaches 80% the second fastest. go/went belongs to the g5 -> wEnt class, which only has a type frequency of 3, so it takes the longest to reach 80% accuracy. In essence, higher type frequency allows verbs to reach 80% accuracy faster.

Q2

Notice that the acquisition order is not directly related to the strength of the scales for each of these words (e.g. higher scales do not mean higher accuracy). Why does this happen? Specifically, what aspect of the model determines the acquisition order of these three verbs? And, how is this aspect of the model sensitive to type frequency?

Type frequency determines the acquisition order of these three verbs– higher type frequency for a class results in faster acquisition of verbs that belong to that class. Since REG verbs are seen the most, the REG class is preferred by the general constraints. Even if set/set and go/went have stronger scales than use/used, use/used is still acquired faster since the REG constraint is preferred by the model, giving it a boost even without as strong of a scale. On the other hand, verbs of the NULL and g5 -> wEnt classes must solely rely on their scale strength to reach higher accuracies, without the headstart from the general constraints.

Token Frequency

All else being equal, children tend to acquire high frequency verbs earlier than lower frequency verbs. To see if the model captures this behavior, you will compare high frequency verbs in one class to mid frequency verbs in the same class. Change the example words to add 'know/knew' (2038), 'grow/grew' (316), and 'buy/bought' (316), in addition to the already included 'think/thought' (4426). Then run the model a few times to answer the following questions.

Q3

Explain whether the model correctly predicts that ‘think/thought’ is acquired before ‘buy/bought’.

The model correctly predicts that ‘think/thought’ is acquired before ‘buy/bought’. This is because the type frequency for ‘think/thought’ is much higher than ‘buy/bought’, giving the model more opportunity to update and strengthen the scale for ‘think/thought’. Therefore, the model does indeed capture the acquisition pattern of children tending to acquire high frequency verbs earlier than lower frequency verbs.

Q4

Explain whether the model correctly predicts that ‘know/knew’ is acquired before ‘grow/grew’.

The model correctly predicts that ‘know/knew’ is acquired before ‘grow/grew’. This is because the type frequency for ‘know/knew’ is much higher than ‘grow/grew’, giving the model more opportunity to update and strengthen the scale for ‘know/knew’. Therefore, the model does indeed capture the acquisition pattern of children tending to acquire high frequency verbs earlier than lower frequency verbs.

EXTRA CREDIT

Explain what predictions the original model makes and how they don’t align with the U-shape learning observations. Then explain what modifications you made, what you hoped they would do, and discuss whether this improved the model’s predictions.

The original model makes predictions that improve over the course of many iterations as it learns each verb in the training data. This does not align with the U-shape learning observations in which children initially produce some high frequency irregular past tense verbs, then arguably regress during the overgeneralization stage, and finally learn exceptions to the regular rule.
For modifications (commented with “extra credit”), I tried:

1. Setting the initial weight of the NULL constraint higher than the others
    a. I hoped that this would reflect the initial stage in children’s language acquisition, in which they do not produce any tense marking at all.
    b. This did indeed reflect language acquisition in children, as for the first few hundred iterations, the NULL constraint was weighted higher than REG, though the model quickly learned to prefer the REG constraint.
2. Making the learning rate for scale updates bigger than the learning rate for weight updates.
    a. I hoped that this would improve the model by better reflecting how quickly children learn rules for specific words.
    b. This did improve the model, as it more quickly learned the rules for irregular verbs, whereas with the original, these verbs were slower to reach high accuracy. For example, g5->wEnt reached 88% accuracy by iteration 49900 with this new model, whereas it took 61900 iterations to reach 80% accuracy with the original model. Additionally, sEt->sEt reached 86% accuracy by iteration 49900 with this new model, whereas it took 69900 iterations to reach 80% accuracy with the original model.