Natural Language Processing Homework 6

Katie Chang

April 14, 2016

README

1 Q1

1.1 a

i. Initially, the probability that day 1 is H was 0.871. With the change, the probability is now 0.491.

ii. With this change, the probability that day 2 is hot is 0.918. This shows a decrease in probability of 0.059. We look at the $p(\to H)$ column for this information.

iii. Originally with two ice creams on day one, after 10 iterations the probability of the first day being hot decreased ever so slightly, but the probability was always around 1.0 (specifically, p(H) = 1.0 and 0.995 on day 1 and 2, respectively)

After the change to one ice cream on day one, however, the data shows that p(H) in general has decreased significantly. Day 1 p(H) is now 0, and at day 2, p(H) = 0.557.

1.2 b

i. Before the change, the graph shows that days 10-13 had a relatively high chance of being hot.

Day 10: 0.918 Day 11: 0.752 Day 12: 0.856 Day 13: 0.779

After this change, we see that the probability of these days being hot decreases quite a bit.

Day 10: 0.764 (doesn't decrease too much because 3 ice creams were eaten) Day 11: 0 (only one ice cream was eaten) Day 12: 0.441 Day 13: 0.441

Since we're restricting our model such that it can never be a hot day if only 1 ice cream was eaten, then the probability of it being hot on day 11 is 0, which affects everything that comes after it.

ii. After 10 iterations, the probability of it being hot on days 12 and 13 drop to nearly 0.

Before the restriction we made of $p(1|H) = 0$, the graph actually looks pretty much the same as it does after we added this restriction. For day 12, the difference of p(H) from before the change to after is a drop of 0.001.

This is because after 10 iterations, this model should learn that the probability of one ice cream given that the day is hot is very low, regardless of our initial restriction. In fact, without this restriction, at the 10th iteration $p(1|H) = 1.6\text{E-}04$, which is basically 0.

iii. With the change of $p(1|H) = 0$, it still equals 0 after 10 iterations. Based on the excel, $p(1|H) = p(-> H, 1) \ / \ p(-> H)$ from the previous iteration. But if we initially had set $p(1|H) = 0$, then $p(-> H, 1)$ is also always 0, so this value is always 0.

1.3   c

i. Since the backward algorithm is analogous to the inside algorithm and the inside algorithm finds the probability of a sentence by summing over all possible parses, this should mean that the β probability of the EOS character sums all possible parses, therefore equaling the total probability of the sentence.

ii. In the tree on the left, the H constituent means that the day is hot. The specific day corresponds to the depth at which the H is at in the tree. The probability of the rule H -¿ 1 C is equal to $p(1|C)$ * $p(C|H)$. The probability of H -¿ ϵ is equal to $p(STOP|H)$. The tree on the right may be preferred because instead of having to calculate p(H -¿ 1 C) as stated above, we can instead get p(H -¿ EC C) from $p(C|H)$, which is already in our excel spreadsheet. Also, p(EC -¿ 1) = $p(1|C)$, which is also in the spreadsheet.

2 vtag

3 vtag

Used overleaf.com to generate LaTeX document.