Prof. Ryan Cotterell

Simon Wachter: Assignment 4

siwachte@ethz.ch, 19-920-198

14/01/2023 - 15:06h

Question 1:

a)

$$
\begin{aligned}
\sum_{w\in\Sigma^*}\tilde{p}(w) &= \sum_{w\in\Sigma^*,|w|=1}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (1)\\
&= \sum_{w\in\Sigma}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (2)\\
&= 1 + \sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (3)\\
&= 1 + \sum_{w\in\Sigma,|w|=1}\sum_{w'\in\Sigma^*}\tilde{p}(w)\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (4)\\
&= 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w'\in\Sigma^*,|w|=1}\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (5)\\
&= 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w'\in\Sigma}\tilde{p}(w') + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (6)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma^*,|w|=3}\tilde{p}(w) + \dots && (7)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma}\sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w)\tilde{p}(w') + \dots && (8)\\
&= 1 + 1\cdot 1 + \sum_{w\in\Sigma}\tilde{p}(w)\sum_{w\in\Sigma^*,|w|=2}\tilde{p}(w') + \dots && (9)\\
&= 1 + 1\cdot 1 + 1\cdot 1\cdot 1 + \dots && (10)\\
&= \sum_{i=1}^{\infty} 1^i && (11)
\end{aligned}
$$

We can see that the series in equation 11 diverges to ∞.

b) We first state some equations that are used later:

$$\sum_{w \in \Sigma \cup \{\text{EOS}\}} p(w) = 1 \tag{12}$$

$$\Rightarrow \sum_{w \in \Sigma} p(w) = 1 - p(\text{EOS}) \tag{13}$$

$$\sum_{w \in \Sigma} p(w) < 1 \tag{14}$$

$$\sum_{w \in \Sigma} p(\text{EOS}) p(w) = p(\text{EOS}) \sum_{w \in \Sigma} p(w) \tag{15}$$

$$\sum_{w \in \Sigma^*, |w|=0} p(w) = p(\text{EOS}) \qquad \text{because we have only one EOS symbol} \tag{16}$$

$$\tag{17}$$

We define l as:

$$l = \sum_{w \in \Sigma} p(w) \tag{18}$$

$$= (1 - p(\text{EOS})) \tag{19}$$

$$l < 1 \tag{20}$$

Now we can formulate the given equation:

$$\sum_{w\in\Sigma^*} p(w) = \sum_{w\in\Sigma^*,|w|=0} p(w) + \sum_{w\in\Sigma^*,|w|=1} p(w) + \sum_{w\in\Sigma^*,|w|=2} p(w) + \sum_{w\in\Sigma^*,|w|=3} p(w) + \dots \tag{21}$$

$$= p(\mathrm{EOS}) + \sum_{w\in\Sigma^*,|w|=1} p(w) + \sum_{w\in\Sigma^*,|w|=2} p(w) + \sum_{w\in\Sigma^*,|w|=3} p(w) + \dots \tag{22}$$

$$= p(\mathrm{EOS}) \left(1 + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{23}$$

$$= p(\mathrm{EOS}) \left(1 + \sum_{w\in\Sigma^*,|w|=1} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{24}$$

$$= p(\mathrm{EOS}) \left(1 + \sum_{w\in\Sigma} p(w) + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{25}$$

$$= p(\mathrm{EOS}) \left(1 + l + \sum_{w\in\Sigma^*,|w|=2} \tilde{p}(w) + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{26}$$

$$= p(\mathrm{EOS}) \left(1 + l + \sum_{w\in\Sigma} \sum_{w'\in\Sigma} p(w)p(w') + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{27}$$

$$= p(\mathrm{EOS}) \left(1 + l + \sum_{w\in\Sigma} p(w) \sum_{w'\in\Sigma} p(w') + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{28}$$

$$= p(\mathrm{EOS}) \left(1 + l + l^2 + \sum_{w\in\Sigma^*,|w|=3} \tilde{p}(w) + \dots\right) \tag{29}$$

$$= p(\mathrm{EOS}) \left(1 + l + l^2 + l^3 + \dots\right) \tag{30}$$

$$= p(\mathrm{EOS}) \left(\sum_{n=0}^{\infty} l^n\right) \tag{31}$$

$$= p(\mathrm{EOS}) \frac{1}{1-l} \qquad \text{(limit geom. series, because } l < 1\text{)} \tag{32}$$

$$= \frac{p(\mathrm{EOS})}{1-(1-p(\mathrm{EOS}))} \tag{33}$$

$$= \frac{p(\mathrm{EOS})}{p(\mathrm{EOS})} \tag{34}$$

$$= 1 \tag{35}$$

c)

$$\sum_{u\in\Sigma^*} p(wu) = \sum_{u\in\Sigma^*} p(\text{EOS}|wu)p_{pre}(u|w)p_{pre}(w) \qquad \text{by def.} \tag{36}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} p(\text{EOS}|wu)p_{pre}(u|w)\right) \tag{37}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} p(u|w)\right) \qquad \text{def. p(w)} \tag{38}$$

$$= p_{pre}(w)\left(\sum_{u\in\Sigma^*} \frac{p(w|u)p(u)}{p(w)}\right) \qquad \text{Bayes rule} \tag{39}$$

$$= p_{pre}(w)\left(\frac{p(w)}{p(w)}\right) \qquad \text{Bayes rule} \tag{40}$$

$$= p_{pre}(w) \tag{41}$$

d) We use CKY with the $(+,\times)$-semiring. Further we use $\log p$ insted of p for our score function. These operations will lead to $p(w_1 \dots w_n|S)$ being calculated at the top of the tree because we fill the tree according to:

$$\text{s}[i,k,X] = \sum_k \sum_i \sum_j \exp(\text{score}(X \to YZ)) \times \text{s}[i,j,Y] \times \text{s}[j,k,Z] \tag{42}$$

$$= \exp(\log(p(YZ|X))) \times p(w_i \dots w_j|Y) \times p(w_{j+1} \dots w_k|Z) \tag{43}$$

$$= p(YZ|X) \times p(w_i \dots w_j|Y) \times p(w_{j+1} \dots w_k|Z) \tag{44}$$

$$= p(w_i \dots w_k|X) \tag{45}$$

We can the easily multiply $p(w_1 \dots w_n|S)$ by $p(\text{EOS})$ and get the desired result.

e)

$$\sum_{u\in\Sigma^*} p(wu) = \sum_{u\in\Sigma^*} p_{\text{inside}}(wu|S) \tag{46}$$

$$= p(X \overset{*}{\Rightarrow} w_i \dots w_k) \tag{47}$$

$$= p(S \overset{*}{\Rightarrow} wv) \tag{48}$$

$$\tag{49}$$

eq. (48) holds, because summing over all possible suffixes is equivalent to allow all different derivation trees that produce v.

f) Following the hint given in the exercise, we create a matrix $M \in |\mathcal{N}| \times |\mathcal{N}|$ where each entry $M_{X,Y}$ corresponds to the probability of deriving Y from X:

$$M_{Y,X} = p(X \to Y\alpha) \tag{50}$$

We can see that by multiplying in the inside semiring this matrix with itself we get the probability of deriving i from j in two steps and so on. Thus, the kleene star over M

will yield the desired property. To derive the kleene star we use Lehmann's algorithm:

$$M^* = \bigotimes_{k=0}^{\infty} M \tag{51}$$
$$= I + M \bigotimes_{k=0}^{\infty} \tag{52}$$
$$= I + MM^* \tag{53}$$

$$M^* - MM^* = I \tag{54}$$
$$(I - M)M^* = I \tag{55}$$
$$M^* = (I - M)^{-1} \tag{56}$$

As stated in the lecture notes, for eq. (56) to holde we require the largest eigenvalue of M to be smaller than 1. This translates to the condition that all diagonal entries of M have to be < 1. The diagonal entries correspond to the probability of deriving a symbol from itself. This probability has to be smaller than 1 in our case, since otherwise we would not be able to generate any other symbol from it than itself, which would results in infinite sentences.
Then, using eq. (56) we can calculate M^* in $\mathcal{O}(|\mathcal{N}|^3)$. The entries in the matrix now correspond to:

$$M^*_{X,Y} = p[X \stackrel{*}{\Rightarrow} Y\alpha] = p_{\text{lc}}(Y|X) \tag{57}$$

From this matrix we further need to derive the $p_{\text{lc}}(YZ|X)$. We iterate over all possible choices for X, X', Y, Z to derive:

$$p_{\text{lc}}(YZ|X) = \sum_{X' \in \Sigma} p_{\text{lc}}(X'|X) p_{\text{lc}}(Y|X') p_{\text{lc}}(Z|X'Y) \tag{58}$$

Which can be done in $\mathcal{O}(|\mathcal{N}|^4)$.

g)

$$p_{pre}(w_i \dots w_k|X) = \sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} \color{red}{p(X \stackrel{*}{\Rightarrow} YZ\alpha)} \color{green}{p(Y \stackrel{*}{\Rightarrow} w_i \dots w_j)} \color{blue}{p(Z \stackrel{*}{\Rightarrow} w_{j+1} \dots w_k)} \tag{59}$$

$$\color{red}{p(X \stackrel{*}{\Rightarrow} YZ\alpha)} = \text{probability of deriving the two subtrees } Y \text{ and } Z \text{ from } X \tag{60}$$
$$\color{green}{p(Y \stackrel{*}{\Rightarrow} w_i \dots w_j)} = \text{probability of deriving the string } w_i \dots w_j \text{ from } Y \tag{61}$$
$$\color{blue}{p(Z \stackrel{*}{\Rightarrow} w_{j+1} \dots w_k)} = \text{probability of deriving the string } w_{j+1} \dots w_k \text{ from } Z \tag{62}$$
$$\sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} \color{red}{p(X \stackrel{*}{\Rightarrow} YZ\alpha)} \color{green}{p(Y \stackrel{*}{\Rightarrow} w_i \dots w_j)} \color{blue}{p(Z \stackrel{*}{\Rightarrow} w_{j+1} \dots w_k)} \tag{63}$$
$$= \sum_{j=1}^{k-1} \sum_{Y,Z \in \mathcal{N}} p_{\text{lc}}(YZ|X) p_{\text{inside}}(w_i \dots w_j|Y) p_{\text{pre}}(w_{j+1} \dots w_k|Z) \tag{64}$$
$$\tag{65}$$

h) Clarify: We can see that if we have all precomputation done, then the calculation of $p_{pre}(w')$ can be done in $\mathcal{O}(1)$ for a given subsequence w' of w. As $|w| = N$, this has runtime $\mathcal{O}(N)$. Now for the precompution, we first look at $p_{\text{lc}}(YZ|X)$. As shown in exercise f) this can be done in $\mathcal{O}(|\mathcal{N}|^4)$. Equation (15) on the exercise sheet shows that $p_{pre}(w_i \dots w_k|X)$ depends only on $p_{\text{lc}}(YZ|X)$, $p_{inside}(w_i \dots w_j)$ and $p_{\text{pre}(w_{j+1} \dots w_k)}$ for $j < k$. The first of the three we have already calculated. We calculate p_{inside} with the CKY algorithm and save all intermediate values. The size of our grammar is $\mathcal{O}(|\mathcal{N}|^2)$. Thus, CKY runs in $\mathcal{O}(N^3|\mathcal{N}|^2)$ and we need to run it $\mathcal{O}(|\mathcal{N}|)$ times for each $X \in \Sigma$. After these steps we have all our value precomputed in $\mathcal{O}(N^3|\mathcal{N}|^3 + |\mathcal{N}|^4)$. We then invoke the algorithm naively for each subsequence of $w_1 \dots w_i$. This gives us a runtime of $\mathcal{O}(N^4|\mathcal{N}|^3 + N|\mathcal{N}|^4)$.

$$s \tag{66}$$