Prof. Ryan Cotterell

Simon Wachter: Assignment 2

siwachte@ethz.ch, 19-920-198

19/11/2022 - 15:43h

Question 1:

a) Prove that the expectation semiring satisfies the semiring axioms:

- $(\mathbb{R} \times \mathbb{R}, \oplus, \mathbf{0})$ must be a commutative monoid with identity element $\mathbf{0}$:

$$\begin{aligned}
(\langle x, y\rangle \oplus \langle x', y'\rangle) \oplus \langle x'', y''\rangle &= \langle x + x', y + y'\rangle \oplus \langle x'', y''\rangle && (1)\\
&= \langle x + x' + x'', y + y' + y''\rangle && (2)\\
&= \langle x, y\rangle \oplus \langle x' + x'', y' + y''\rangle && (3)\\
&= \langle x, y\rangle \oplus (\langle x', y'\rangle \oplus \langle x'', y''\rangle) && (4)
\end{aligned}$$

$$\begin{aligned}
\mathbf{0} + \langle x, y\rangle &= \langle 0, 0\rangle \oplus \langle x, y\rangle && (5)\\
&= \langle 0 + x, 0 + y\rangle && (6)\\
&= \langle x, y\rangle && (7)\\
&= \langle x + 0, y + 0\rangle && (8)\\
&= \langle x, y\rangle + \mathbf{0} && (9)
\end{aligned}$$

$$\begin{aligned}
\langle x, y\rangle + \langle x', y'\rangle &= \langle x + x', y + y'\rangle && (10)\\
&= \langle x' + x, y' + y\rangle && (11)\\
&= \langle x', y'\rangle + \langle x, y\rangle && (12)
\end{aligned}$$

- $(\mathbb{R} \times \mathbb{R}, \otimes, \mathbf{1})$ must be a monoid with identity element $\mathbf{1}$:

$$\begin{aligned}
(\langle x, y\rangle \otimes \langle x', y'\rangle) \otimes \langle x'', y''\rangle &= \langle x \cdot x', x \cdot y' + y \cdot x'\rangle \otimes \langle x'', y''\rangle && (13)\\
&= \langle x \cdot x' \cdot x'', x \cdot x' \cdot y'' + (x \cdot y' + y \cdot x') \cdot x''\rangle && (14)\\
&= \langle x \cdot x' \cdot x'', x \cdot x' \cdot y'' + x \cdot y' \cdot x'' + y \cdot x' \cdot x''\rangle && (15)\\
&= \langle x, y\rangle \otimes \langle x' \cdot x'', x' \cdot y'' + y' \cdot x''\rangle && (16)\\
&= \langle x, y\rangle \otimes (\langle x', y'\rangle \otimes \langle x'', y''\rangle) && (17)
\end{aligned}$$

$$\begin{aligned}
\mathbf{1} \otimes \langle x, y\rangle &= \langle 1, 0\rangle \otimes \langle x, y\rangle && (18)\\
&= \langle 1 \cdot x, 1 \cdot y\rangle && (19)\\
&= \langle x, y\rangle && (20)\\
&= \langle x \cdot 1, y \cdot 1\rangle && (21)\\
&= \langle x, y\rangle \otimes \mathbf{1} && (22)
\end{aligned}$$

- Multiplication left and right distributes over addition:

$$\begin{aligned}
\langle x, y\rangle \otimes (\langle x', y'\rangle \oplus \langle x'', y''\rangle) &= \langle x, y\rangle \otimes \langle x' + x'', y' + y''\rangle && (23)\\
&= \langle x \cdot x' + x \cdot x'', x \cdot y' + x \cdot y'' + y \cdot x' + y \cdot x''\rangle && (24)\\
&= \langle x \cdot x', x \cdot y' + y \cdot x'\rangle \oplus \langle x \cdot x'', x \cdot y'' + y \cdot x''\rangle && (25)\\
&= (\langle x, y\rangle \otimes \langle x', y'\rangle) \oplus (\langle x, y\rangle \otimes \langle x'', y''\rangle) && (26)
\end{aligned}$$

$$\begin{aligned}
(\langle x, y\rangle \oplus \langle x', y'\rangle) \otimes \langle x'', y''\rangle &= \langle x + x', y + y'\rangle \otimes \langle x'', y''\rangle && (27)\\
&= \langle x \cdot x'' + x' \cdot x'', x \cdot y'' + x' \cdot y'' + y \cdot x'' + y' \cdot x''\rangle && (28)\\
&= \langle x \cdot x'', x \cdot y'' + y \cdot x''\rangle \oplus \langle x' \cdot x'', x' \cdot y'' + y' \cdot x''\rangle && (29)\\
&= (\langle x, y\rangle \otimes \langle x'', y''\rangle) \oplus (\langle x', y'\rangle \otimes \langle x'', y''\rangle) && (30)
\end{aligned}$$

- Multiplication by $\mathbf{0}$ annihilates $\mathbb{R} \times \mathbb{R}$:

$$\begin{aligned}
\mathbf{0} \otimes \langle x, y\rangle &= \langle 0, 0\rangle \otimes \langle x, y\rangle && (31)\\
&= \langle 0 \cdot x, 0 \cdot y\rangle && (32)\\
&= \langle 0, 0\rangle && (33)\\
&= \mathbf{0} && (34)\\
&= \langle 0, 0\rangle && (35)\\
&= \langle x \cdot 0, y \cdot 0\rangle && (36)\\
&= \langle x, y\rangle \otimes \langle 0, 0\rangle && (37)\\
&= \langle x, y\rangle \otimes \mathbf{0} && (38)
\end{aligned}$$

b) Our initial graph looks like **??** .

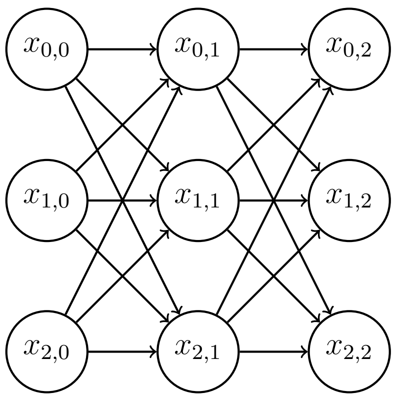

Figure 1: The initial graph

Where the columns represent the words in $\mathbf{w}$ and the rows represent different tags. In this non lifted graph, our forward propagation will yield the following values, when using the score functions as edge weights in the graph:

$$z_{i,j} = \sum_{k=0}^{T} (z_{k,j-1} \cdot \text{score}(x_{k,j-1}, x_{i,j}, \mathbf{w})) \tag{39}$$

Algorithm 1: Forward pass

$\beta(\mathbf{w}, t_0) = 1$
for $n = 1 \to N$ **do**
  $\beta(\mathbf{w}, t_n) = \sum_{t_{n-1} \in \mathcal{T}} \exp(\text{score}(\langle t_n, t_{n-1} \rangle, \mathbf{w})) \otimes \beta(\mathbf{w}, t_{n-1})$
end

Where we used the algorithm from the script: When we now lift the CRF into the expectation semiring, the forward propagation algorithm changes to:

Algorithm 2: Forward pass

$\beta(\mathbf{w}, t_0) = \langle 1, 0 \rangle$
for $n = 1 \to N$ **do**
  $\beta(\mathbf{w}, t_n) = \oplus_{t_{n-1} \in \mathcal{T}} \langle w, -w \log w \rangle \otimes \beta(\mathbf{w}, t_{n-1})$
end

Where $w = \exp(\text{score}(\langle t_n, t_{n+1} \rangle, \mathbf{w}))$.
We want to show that the result of the forward propagation lifted in the semiring is the same as the unnormalized Entropy:

$$H_u(T_w) = -\sum_{\mathbf{t} \in \mathcal{T}^N} \exp(score_{\boldsymbol{\theta}}(\mathbf{t}, \boldsymbol{w})) score_{\boldsymbol{\theta}}(\mathbf{t}, \boldsymbol{w}) \tag{40}$$

$$\tag{41}$$

We show this by induction