Prof. Ryan Cotterell

Yannick Wattenberg: Assignment 02

ywattenberg@inf.ethz.ch, 19-947-464.

12/11/2022 - 12:41h

1 Question: Entropy of a Conditional Random Field

a)

Prove that the **expectation semiring** satisfies the semiring axioms:
Let $x_1, y_1, x_2, y_2, x_3, y_3 \in \mathbb{R}$ for this subsection.

Axiom 1

$$(\mathbb{R} \times \mathbb{R}, \oplus, \langle 0, 0\rangle) \text{ is a commutative monoid with identity element } \langle 0, 0\rangle$$

Associativity and Commutativity of $\oplus$:

$$\begin{aligned}
(\langle x_1, y_1\rangle \oplus \langle x_2, y_2\rangle) \oplus \langle x_3, y_3\rangle &\overset{!}{=} \langle x_1, y_1\rangle \oplus (\langle x_3, y_3\rangle \oplus \langle x_2, y_2\rangle) && (1)\\
(\langle x_1, y_1\rangle \oplus \langle x_2, y_2\rangle) \oplus \langle x_3, y_3\rangle &= \langle x_1 + x_2, y_1 + y_2\rangle \oplus \langle x_3, y_3\rangle && \text{(def. } \oplus\text{)} \quad (2)\\
&= \langle (x_1 + x_2) + x_3, (y_1 + y_2) + y_3\rangle && \text{(def. } \oplus\text{)} \quad (3)\\
&= \langle x_1 + (x_2 + x_3), y_1 + (y_2 + y_3)\rangle && \text{(ass. of +)} \quad (4)\\
&= \langle x_1, y_1\rangle \oplus \langle x_2 + x_3, y_2 + y_3\rangle && \text{(def. } \oplus\text{)} \quad (5)\\
&= \langle x_1, y_1\rangle \oplus \langle x_3 + x_2, y_3 + y_2\rangle && \text{(comm. of +)} \quad (6)\\
&= \langle x_1, y_1\rangle \oplus (\langle x_3, y_3\rangle \oplus \langle x_2, y_2\rangle) && \text{(def. } \oplus\text{)} \quad (7)
\end{aligned}$$

$\langle 0, 0\rangle$ is the identity element:

$$\begin{aligned}
\langle 0, 0\rangle \oplus \langle x_1, y_1\rangle &\overset{!}{=} \langle x_1, y_1\rangle && (8)\\
\langle 0, 0\rangle \oplus \langle x_1, y_1\rangle &= \langle x_1 + 0, y_1 + 0\rangle && \text{(def. } \oplus\text{)} \quad (9)\\
&= \langle x_1, y_1\rangle \quad = \langle x_1 + 0, y_1 + 0\rangle = \langle x_1, y_1\rangle \oplus \langle 0, 0\rangle && \text{(def. } \oplus\text{)} \quad (10)
\end{aligned}$$

Axiom 2

$$(\mathbb{R} \times \mathbb{R}, \otimes, \langle 1, 0\rangle) \text{ is a monoid with identity element } \langle 1, 0\rangle$$

Associativity of $\otimes$:

$$\begin{aligned}
&(\langle x_1, y_1\rangle \otimes \langle x_2, y_2\rangle) \otimes \langle x_3, y_3\rangle \stackrel{!}{=} \langle x_1, y_1\rangle \otimes (\langle x_2, y_2\rangle \otimes \langle x_3, y_3\rangle) && && (11)\\
&(\langle x_1, y_1\rangle \otimes \langle x_2, y_2\rangle) \otimes \langle x_3, y_3\rangle = \langle x_1 \cdot x_2, x_1 \cdot y_2 + y_1 \cdot x_2\rangle \otimes \langle x_3, y_3\rangle && \text{(def. } \otimes) && (12)\\
=&\langle (x_1 \cdot x_2) \cdot x_3, (x_1 \cdot x_2) \cdot y_3 + (x_1 \cdot y_2 + y_1 \cdot x_2) \cdot x_3\rangle && \text{(def. } \otimes) && (13)\\
=&\langle (x_1 \cdot x_2) \cdot x_3, x_1 \cdot x_2 \cdot y_3 + x_1 \cdot y_2 \cdot x_3 + y_1 \cdot x_2 \cdot x_3\rangle && \text{(diss. } \cdot) && (14)\\
=&\langle (x_1 \cdot x_2) \cdot x_3, x_1 \cdot (x_2 \cdot y_3 + y_2 \cdot x_3) + y_1 \cdot x_2 \cdot x_3\rangle && \text{(diss. } \cdot) && (15)\\
=&\langle x_1 \cdot (x_2 \cdot x_3), x_1 \cdot (x_2 \cdot y_3 + y_2 \cdot x_3) + y_1 \cdot (x_2 \cdot x_3)\rangle && \text{(ass. } \cdot) && (16)\\
=&\langle x_1, y_1\rangle \otimes \langle x_2 \cdot x_3, x_2 \cdot y_3 + y_2 \cdot x_3\rangle && \text{(def. } \otimes) && (17)\\
=&\langle x_1, y_1\rangle \otimes (\langle x_2, y_2\rangle \otimes \langle x_3, y_3\rangle) && \text{(def. } \otimes) && (18)
\end{aligned}$$

$\langle 1, 0\rangle$ is the identity element:

$$\begin{aligned}
\langle x_1, y_1\rangle \otimes \langle 1, 0\rangle &\stackrel{!}{=} \langle x_1, y_1\rangle && && (19)\\
\langle x_1, y_1\rangle \otimes \langle 1, 0\rangle &= \langle x_1 \cdot 1, x_1 \cdot 0 + y_1 \cdot 1\rangle && \text{(def. } \otimes) && (20)\\
&= \langle x_1, y_1\rangle \qquad = \langle 1 \cdot x_1, 1 \cdot y_1 + 0 \cdot x_1\rangle && && (21)\\
&= \langle 1, 0\rangle \otimes \langle x_1, y_1\rangle && \text{(def. } \otimes) && (22)
\end{aligned}$$

Axiom 3

$\otimes$ distributes left and right over $\oplus$:

$$\langle x_1, y_1\rangle \otimes (\langle x_2, y_2\rangle \oplus \langle x_3, y_3\rangle) \stackrel{!}{=} (\langle x_1, y_1\rangle \otimes \langle x_2, y_2\rangle) \oplus (\langle x_1, y_1\rangle \otimes \langle x_3, y_3\rangle) \qquad (23)$$

$$\begin{aligned}
&\langle x_1, y_1\rangle \otimes (\langle x_2, y_2\rangle \oplus \langle x_3, y_3\rangle) = \langle x_1, y_1\rangle \otimes \langle x_2 + x_3, y_2 + y_3\rangle && \text{(def. } \oplus) && (24)\\
=&\langle x_1 \cdot (x_2 + x_3), x_1 \cdot (x_2 + x_3) + y_1 \cdot (y_2 + y_3)\rangle && \text{(def. } \otimes) && (25)\\
=&\langle (x_1 \cdot x_2) + (x_1 \cdot x_3), (x_1 \cdot y_2) + (y_1 \cdot x_2) + (x_1 \cdot y_3) + (y_1 \cdot x_3)\rangle && \text{(diss. } \cdot) && (26)\\
=&\langle x_1 \cdot x_2, (x_1 \cdot y_2) + (y_1 \cdot x_2)\rangle \oplus \langle x_1 \cdot x_3, (x_1 \cdot y_3) + (y_1 \cdot x_3)\rangle && \text{(def. } \oplus) && (27)\\
=&(\langle x_1, y_1\rangle \otimes \langle x_2, y_2\rangle) \oplus (\langle x_1, y_1\rangle \otimes \langle x_3, y_3\rangle) && \text{(def. } \otimes) && (28)
\end{aligned}$$

$$(\langle x_2, y_2\rangle \oplus \langle x_3, y_3\rangle) \otimes \langle x_1, y_1\rangle \stackrel{!}{=} (\langle x_2, y_2\rangle \otimes \langle x_1, y_1\rangle) \oplus (\langle x_3, y_3\rangle \otimes \langle x_1, y_1\rangle) \qquad (29)$$

$$\begin{aligned}
&(\langle x_2, y_2\rangle \oplus \langle x_3, y_3\rangle) \otimes \langle x_1, y_1\rangle = (\langle x_2 + x_3, y_2 + y_3\rangle) \otimes \langle x_1, y_1\rangle && \text{(def. } \oplus) && (30)\\
=&\langle (x_2 + x_3) \cdot x_1, (x_2 + x_3) \cdot y_1 + (y_2 + y_3) \cdot x_1\rangle && \text{(def. } \otimes) && (31)\\
=&\langle (x_2 \cdot x_1) \cdot (x_3 \cdot x_1), (x_2 \cdot y_1) + (x_3 \cdot y_1) + (y_2 \cdot x_1) + (y_3 \cdot x_1)\rangle && \text{(diss. } \cdot) && (32)\\
=&\langle x_2 \cdot x_1, x_2 \cdot y_1 + y_2 \cdot x_1\rangle \oplus \langle x_3 \cdot x_1, x_3 \cdot y_1 + y_3 \cdot x_1\rangle && \text{(def. } \oplus) && (33)\\
=&(\langle x_2, y_2\rangle \otimes \langle x_1, y_1\rangle) \oplus (\langle x_3, y_3\rangle \otimes \langle x_1, y_1\rangle) && \text{(def. } \otimes) && (34)
\end{aligned}$$

Axiom 4

$$\langle 0,0\rangle \otimes \langle x_1,y_1\rangle \stackrel{!}{=} \langle 0,0\rangle \stackrel{!}{=} \langle x_1,y_1\rangle \otimes \langle 0,0\rangle \tag{35}$$
$$\langle 0,0\rangle \otimes \langle x_1,y_1\rangle = \langle 0\cdot x_1, 0\cdot y_1 + 0\cdot x_1\rangle \quad \text{(def. } \otimes) \tag{36}$$
$$= \langle 0,0\rangle \tag{37}$$
$$= \langle x_1\cdot 0, x_1\cdot 0 + y_1\cdot 0\rangle \tag{38}$$
$$= \langle x_1,y_1\rangle \otimes \langle 0,0\rangle \quad \text{(def. } \otimes) \tag{39}$$

With this we have proven that the **expectation semiring** is a semiring.

b)

The definition of the forward algorithm is as follows: We raise this into the expectation

---

Algorithm 1: Forward algorithm

$\beta(\mathbf{w}, t_0) = \mathbb{1}$
for $i = 1$ **to** N **do**
  $\beta(\mathbf{w}, t_i) = \oplus_{t_i \in T} exp(score(t_i, t_{i+1}, \mathbf{w})) \otimes \beta(\mathbf{w}, t_{i-1})$
end

---

semiring by replacing the $\oplus$ and $\otimes$ with the corresponding operations in the expectation semiring. We also replace the initialization of $\beta(w, t_0)$ with the neutral element of multiplication in the expectation semiring. Which yields the following algorithm:

Algorithm 2: Forward algorithm in the expectation semiring

$\beta(\mathbf{w}, t_0) = \langle 1, 0\rangle$
for $i = 1$ **to** N **do**
  $w = exp(score(t_{i-1}, t_i, \mathbf{w}))$
  $\beta(\mathbf{w}, t_i) = \oplus_{t_i \in T} \langle w, -w \cdot logw\rangle \otimes \beta(\mathbf{w}, t_{i-1})$
end

We want to prove that the result of the forward algorithm lifted in the expectation semiring computes the unnormalized Entropy defined as:

$$H_u(T_{\mathbf{w}}) = -\sum_{t\in T^N} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w}) \tag{40}$$
$$\tag{41}$$

We prove the statement by induction on the length of the sequence N. The base case is trivial, since the forward algorithm only computes the score of the first token:

$$\beta(\mathbf{w}, t_1) = \oplus_{t_1\in T} \langle w, -w\cdot logw\rangle \otimes \langle 1, 0\rangle \tag{42}$$
$$\text{(def. } \otimes) = \oplus_{t_1\in T} \langle w, -w\cdot logw\rangle \tag{43}$$
$$\text{(def. } w) = \oplus_{t_1\in T} \langle exp(score(t_0, t_1, \mathbf{w})), -exp(score(t_0, t_1, \mathbf{w}))\cdot score(t_0, t_1, \mathbf{w})\rangle \tag{44}$$
$$\text{(def. } \oplus) = \langle \sum_{t\in T^1} exp(score(t_0, t_1, \mathbf{w})), -\sum_{t\in T^1} exp(score_\Theta(t, \mathbf{w}))\cdot score_\Theta(t, \mathbf{w})\rangle \tag{45}$$

Our induction hypothesis is that the forward algorithm computes the unnormalized entropy for sequences of length i:

$$\beta(\mathbf{w}, t_i) = \langle \sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})), -\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w}) \rangle \tag{46}$$

Induction step ($i \rightarrow i+1$):

$$\begin{aligned}
\beta(\mathbf{w}, t_{i+1}) &= \oplus_{t_{i+1} \in T} \langle w, -w \cdot logw \rangle \otimes \beta(\mathbf{w}, t_i) && (47)\\
&= \oplus_{t_{i+1} \in T} \langle w, -w \cdot logw \rangle \otimes && (48)\\
&\quad \langle \sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})), -\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w}) \rangle \quad \text{(IH)} && (49)\\
&= \langle \sum_{t_{i+1} \in T} \sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})), && (50)\\
&\quad \sum_{t_{i+1} \in T} w \cdot (-\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w})) && (51)\\
&\quad - \sum_{t_{i+1} \in T} w \cdot log(w) \cdot (\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w}))) \rangle && (52)\\
(50) \Rightarrow & \sum_{t_{i+1} \in T} w \cdot \sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) && (53)\\
(\text{def. } w) &= \sum_{t_{i+1} \in T} exp(score(t_i, t_{i+1}, \mathbf{w})) \cdot \sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) && (54)\\
&= \sum_{t \in T^i} \sum_{t_{i+1} \in T} exp(score_\Theta(t, \mathbf{w}) + score(t_i, t_{i+1} \mathbf{w})) && (55)\\
&= \sum_{t \in T^{i+1}} exp(score_\Theta(t, \mathbf{w})) && (56)\\
(51) \Rightarrow & \sum_{t_{i+1} \in T} w \cdot (-\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w})) && (57)\\
(\text{def. } w) &= \sum_{t_{i+1} \in T} exp(score(t_i, t_{i+1}, \mathbf{w})) \cdot (-\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w})) && (58)\\
&= -\sum_{t \in T^i} \sum_{t_{i+1} \in T} exp(score_\Theta(t, \mathbf{w}) + score(t_i, t_{i+1}, \mathbf{w})) \cdot score_\Theta(t, \mathbf{w}) && (59)\\
(52) \Rightarrow & \sum_{t_{i+1} \in T} w \cdot log(w) \cdot (\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w}))) && (60)\\
(\text{def. } w) &= \sum_{t_{i+1} \in T} exp(score(t_i, t_{i+1}, w)) \cdot score(t_i, t_{i+1}, w) \cdot (\sum_{t \in T^i} exp(score_\Theta(t, \mathbf{w}))) && (61)\\
&= \sum_{t \in T^i} \sum_{t_{i+1} \in T} exp(score_\Theta(t, \mathbf{w}) + score(t_i, t_{i+1}, \mathbf{w})) \cdot score(t_i, t_{i+1}, \mathbf{w}) && (62)
\end{aligned}$$

$$(59-62) \Rightarrow -\sum_{t\in T^i}\sum_{t_{i+1}\in T} exp(score_\Theta(t,\mathbf{w}) + score(t_i,t_{i+1},\mathbf{w}))\cdot score_\Theta(t,\mathbf{w}) \tag{63}$$

$$-\sum_{t\in T^i}\sum_{t_{i+1}\in T} exp(score_\Theta(t,\mathbf{w}) + score(t_i,t_{i+1},\mathbf{w}))\cdot score(t_i,t_{i+1},\mathbf{w}) \tag{64}$$

$$= -\sum_{t\in T^i}\sum_{t_{i+1}\in T} exp(score_\Theta(t,\mathbf{w}) + score(t_i,t_{i+1},\mathbf{w}))\cdot (score_\Theta(t,\mathbf{w})\cdot score(t_i,t_{i+1},\mathbf{w})) \tag{65}$$

$$= -\sum_{t\in T^{i+1}} exp(score_\Theta(t,\mathbf{w}))\cdot (score_\Theta(t,\mathbf{w})) \tag{66}$$

Together (56) and (66), we have:

$$\langle \sum_{t\in T^{i+1}} exp(score_\Theta(t,\mathbf{w})), \sum_{t\in T^{i+1}} exp(score_\Theta(t,\mathbf{w}))\cdot score_\Theta(t,\mathbf{w})\rangle \tag{67}$$

Concluding the proof. From this we can follow that

$$\beta(w,t_N) = \langle \sum_{t\in T^N} exp(score_\Theta(t,\mathbf{w})), -\sum_{t\in T^N} exp(score_\Theta(t,\mathbf{w}))\cdot score_\Theta(t,\mathbf{w})\rangle \tag{68}$$

Which shows that the second part of the pair is equal to the unnormalized entropy.

c)