Yiwen Zhang

yz3310

COMS 4705 Hw1.

Question 1:

$$perplexity = 2^{-l} \quad where \quad l = \frac{1}{M} \sum_{i=1}^{n} \log_2 P(x^{(i)})$$

$$\underset{\lambda_1,\lambda_2,\lambda_3}{\arg\min}\, 2^{-l} = \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max}\, l = \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max}\, \log_2 \prod_{i=1}^{n} P(x^{(i)})$$

$$= \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max}\, \log_2 \left[\prod_{i=1}^{n}\prod_{j=1}^{m} q(x_j^{(i)} \mid x_{j-2}^{(i)}, x_{j-1}^{(i)})\right]$$

(m is the number of words in ~~each~~ each sentence)

The format of $x_{j-2}^{(i)}, x_{j-1}^{(i)}, x_j^{(i)}$ equivalent to sequence w_1, w_2, w_3.

$C'(w_1, w_2, w_3)$ is the number of times the triagram seen in validation set.

$$\therefore \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\min}\, 2^{-l} = \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max}\, \log_2 \left[\prod_{w_1,w_2,w_3} [q(w_3 \mid w_1, w_2)]^{C'(w_1,w_2,w_3)}\right]$$

$$= \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max} \sum_{w_1,w_2,w_3} C'(w_1, w_2, w_3) \cdot \log_2 q(w_3 \mid w_1, w_2)$$

$$= \underset{\lambda_1,\lambda_2,\lambda_3}{\arg\max}\, L(\lambda_1, \lambda_2, \lambda_3)$$

$\therefore$ ~~λ values~~ λ_i for $i=1,2,3$ values maximize $L(\lambda_1, \lambda_2, \lambda_3)$ also ~~max~~ minimize perplexity.

Question 2:

In the lecture slides, we have $q(w|u,v) = \lambda_1 \times q_{ML}(w|u,v) + \lambda_2 \times q_{ML}(w|v) + \lambda_3 \times q_{ML}(w)$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$

under this definition, the estimate correctly defines a distribution:

$$\sum_{w \in V'} q(w|u,v) = \sum_{w \in V'} [\lambda_1 \times q_{ML}(w|u,v) + \lambda_2 \times q_{ML}(w|v) + \lambda_3 \times q_{ML}(w)]$$

$$= \lambda_1 \times \sum_{w \in V'} q_{ML}(w|u,v) + \lambda_2 \times \sum_{w \in V'} q_{ML}(w|v) + \lambda_3 \times \sum_{w \in V'} q_{ML}(w)$$

$$= \lambda_1 \times 1 + \lambda_2 \times 1 + \lambda_3 \times 1$$

$$= \lambda_1 + \lambda_2 + \lambda_3 = 1$$

In the definition of this question, $\sum_{w \in V'} q(w|u,v) = \sum_{w \in V'} [\lambda_1^{\phi(u,v,w)} \times q_{ML}(w|u,v)$

$+ \lambda_2^{\phi(u,v,w)} \times q_{ML}(w|v)$

$+ \lambda_3^{\phi(u,v,w)} \times q_{ML}(w)]$

$$= \sum_{w \in V'} \lambda_1^{\phi(u,v,w)} \times q_{ML}(w|u,v) + \sum_{w \in V'} \lambda_2^{\phi(u,v,w)} \times q_{ML}(w|v)$$

$$+ \sum_{w \in V'} \lambda_3^{\phi(u,v,w)} \times q_{ML}(w)$$

$\because$ $\lambda_i^{\phi(u,v,w)}$ is not constant, which depends on w. $(i=1,2,3)$

$\therefore$ $\lambda_i^{\phi(u,v,w)}$ $(i=1,2,3)$ cannot be pulled out from the equation like the process in the lecture slides.

$\therefore$ It cannot ensure the sum is still one.

Question 3: Input: a sentence $x_1 \cdots x_n$, parameters $q(s|u,v)$ and $e(x|s)$

Initialization: Set $\pi(0,*,*) = 1$

Definition: $S_{-1} = S_0 = \{*\}$, $S_k = T(x_k)$ for $k \in \{1, \cdots n\}$

Algorithm:

for $k = 1 \ldots n$,

for $u \in S_{k-1}$, $v \in S_k$,

$$\pi(k,u,v) = \max_{w \in S_{k-2}} (\pi(k-1,w,u) \times q(v|w,u) \times e(x_k|v))$$

Return $\max_{u \in S_{n-1}, v \in S_n} (\pi(n,u,v) \times q(STOP|u,v))$