Alexis Tabin
16 - 821 - 803

Computational Semantics
for Natural Language Processing

Assignment 1

2. Multi-prototype Word Embeddings

- (w, c), pair of a word and a context
- N_w, word senses (prototypes
- $W \in \mathbb{R}^{|V_w| \times N_w \times d}$
- $h_w \in \{1, \dots, N_w\}$, $h_w = h$ ~~means~~ if w means if h^{th} prototype
- $\pi_{wh} = P(h_w = h | w)$

We model $$P(D=1 | w, c) = \sum_{h=1}^{N_w} P(h_w = h | w) P(D=1 | w, h_w = h, c)$$

$$= \sum_{h=1}^{N_w} \pi_{wh} P(D=1 | w, h_w = h, c)$$

2.1. $P(D)$, $L(D) = ?$

likelihood, log-likelihood

$$P(D) = \prod_{(w,c) \in D} P_r [D = 1 | w, c]$$

$$= \prod_{(w,c) \in D} \sum_{h=1}^{N_w} P(h_w = h | w) P(D=1 | w, h_w = h, c)$$

$$L(D) = \log(P(D)) = \log\left(\prod_{(w,c)} \sum_{h=1}^{N_w} \pi_{wh} P(D=1 | w, h_w = h, c)\right)$$

$$= \sum_{(w,c) \in D} \log\left(\sum_{h=1}^{N_w} \pi_{wh} P(D=1 | w, h_w = h, c)\right)$$

2.2 $M_{whc} \in \{0,1\} := \mathbb{1}\{h_w = h\}$

$P(D, M)$, $\log(P(D,M)) = L(D,M)$

$\triangleright\ P(D, M) = \underline{P(M|D)}\ \underline{P(D)}$

$P(M|D) = P(M=0\ |D)\,P(M=0) + P(M=1\,|D)\,P(M=1)$

D
$$\prod_{(w,c) \in D} \sum_{h=1}^{N_w} \pi_{wh}\, P(D=1\,|\,w, h_w = h, c) \cdot P(M|D)$$

$$= \prod_{(w,c) \in D} \prod_{h=1}^{N_w} \pi_{wh}\, P(D=1\,|\,w, h_w = h, c)^{M_{whc}}$$

$L(D,M) =$

$$\triangleright\ \log(P(D,M)) = \log \prod_{(w,c)} \prod_{h=1}^{N_w} \pi_{wh}\, P(D=1\,|\,w, h_w=h, c)^{M_{whc}}$$

$$= \sum_{(w,c)} \sum_{h=1}^{N_w} M_{whc} \log(P(D=1\,|\,w, h_w=h, c) + \log(\pi_{wh})$$

2.3 E-step : $Q(\Theta) := \mathbb{E}_{M|D}\, L(D,M)$

$$\mathbb{E}_{M|D} M_{whc} = \mu_{whc}$$

$$\triangleright\ Q(\Theta) = \mathbb{E}_{M|D}\, L(D,M) = \mathbb{E}\left[\sum_{(w,c)\in D} \sum_{h=1}^{N_w} M_{whc} \log(P(D=1|\ldots)) + \log(\pi_{wh}\right.$$

$$= \sum_{(w,c)\in D} \sum_{h=1}^{N_w} \mathbb{E}_{M|D} M_{whc} \log\left(P(D=1|w, h_w=h, c) + \log(\pi_{wh})\right)$$

$$= \sum_{(w,c)\in D} \sum_{h=1}^{N_w} \mu_{whc} \log\left(P(D=1|w, h_w=h, c) + \log(\pi_{wh})\right)$$

parameters Θ : π_{wh} for $h=1$ to N_w

~~and μ_{whc}~~

~~Solution~~ 2.4.

parameters Θ are $\left\{\pi_{wh'}\right\}_{h'=1}^{N_w}$

2.4 write μ_{whc} as $\sim \{\pi_{wh'}\}_{h'=1}^{N_w}$

and $P(\underset{=1}{D} \mid w, h_w = h', c)$

$$\mu_{whc} = \mathbb{E}_{M|D} M_{whc} = 1 \cdot P(M = 1 \mid D) \cancel{+ 0 \cdot P(M = 0 \mid D)}$$

$M = 1$ if $h_w = h$ (2.2)

$$= P(h_w = h \mid D, w) = \frac{P(D \mid w, h_w = h)\underbrace{P(h_w = h \mid w)}_{\pi_{wh}}}{P(D \mid w)}$$

$$= \frac{\pi_{wh} P(D \mid w, h_w = h)}{P(D \mid w)}$$

for $D = 1$, we have : $\mu_{whc} = \dfrac{\pi_{wh} P(D \mid w, h_w = h)}{\sum_{h=1}^{N_w} P(D = 1 \mid w, h_w = h, c)}$

2-5 | M-step

▷ $\pi^{*}_{wh} := \arg\max_{\pi_{wh}} Q(\Theta)$ parameter

$$Q(\Theta) \overset{2.3}{\overset{!}{=}} \sum_{(w,c)\in D} \sum_{h=1}^{N_w} \Big(\underbrace{\mu_{whc}\Big(\log\big(P(D=1 \mid w, h_w = h, c)\big)}_{\text{no } \pi_{whc}}\Big) + \log(\pi_{wh}$$

$$\Rightarrow \arg\max_{\pi_{whc}} Q(\Theta) = \arg\max_{\pi_{whc}} \left(\sum_{(w,c)} \sum^{N_w} \mu_{whc} \log(\pi_{whc}) \right)$$

lagrangian constraint: $\sum^{N_w} \pi_{whc} = 1$

$$\Rightarrow \mathcal{L}(\pi_{whc}; \lambda) = \sum_{(w,c)} \sum^{N_w} \mu_{whc} \log(\pi_{whc}) + \lambda \Big(1 - \sum_{h=1}^{N_w} \pi_{wh}$$

now we derive the lagrangian

$\frac{\partial \mathcal{L}}{\partial \pi_{whc}} = 0 \iff$ and with a few calculations, we have

$\pi^{*}_{whc} = C \cdot \sum_{(w,c)} \mu_{whc}$ where c is just a scaling cst