COMS 4705: Natural Language Processing HW2

Zhuangyu Ren(zr2209)

February 23, 2019

Question 1

1. We know that

$$\frac{\mathrm{d}L(v)}{\mathrm{d}v_k} = \sum_{i=1}^{n} f_k(x_i, y_i) - \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_k(x_i, y') p(y'|x_i; v) - 2Cv_k$$

   As f_1 and f_2 are identical, the value of

$$\sum_{i=1}^{n} f_1(x_i, y_i) - \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_1(x_i, y') p(y'|x_i; v)$$

   and

$$\sum_{i=1}^{n} f_2(x_i, y_i) - \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_2(x_i, y') p(y'|x_i; v)$$

   should be the same.
   In log-linear model, we always want the gradient to be equal to 0, so v_k^* should be equal to

$$\sum_{i=1}^{n} f_k(x_i, y_i) - \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_k(x_i, y') p(y'|x_i; v)$$

   So v_1^* and v_2^* should also be equal as their formula have the same value.

2. Now we have

$$\frac{\mathrm{d}L(v)}{\mathrm{d}v} = \sum_{i=1}^{n} f(x_i, y_i) - \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f(x_i, y') p(y'|x_i; v) - C \cdot sign(v_k)$$

   So $sign(v_1) = sign(v_2)$ or $v_1 = v_2 = 0$

Question 2

$$\frac{\mathrm{d}L(v)}{\mathrm{d}v_k} = \sum_{i=1}^{n} f_k(x_i, y_i) - \sum_{i=1}^{n}\sum_{y' \in \mathcal{Y}} f_k(x_i, y')p(y'|x_i; v)$$

v^* is the optimal solution to the equation above, so

$$\sum_{i=1}^{n} f_k(x_i, y_i) - \sum_{i=1}^{n}\sum_{y' \in \mathcal{Y}} f_k(x_i, y')p(y'|x_i; v^*) = 0$$

For a particular bigram (w_1, w_2),it will only fit feature k, so

$$\sum_{i=1}^{n} f(w_1, w_2) = count(w_1, w_2)$$

Because only when $y = w_2$, $f_k = 1$,otherwise, $f_k = 0$, so

$$\sum_{i=1}^{n}\sum_{y' \in \mathcal{Y}} f_k(x_i, y')p(y'|x_i; v^*) = \sum_{i=1}^{n} f_k(x_i, w_2)p(w_2|x_i; v^*)$$

Only when $x = w_1$, $f_k = 1$, otherwise, $f_k = 0$

$$\sum_{i=1}^{n} f_k(x_i, w_2)p(w_2|x_i; v^*) = \sum_{x=w_1} f_k(w_1, w_2)p(w_2|w_1, v^*) = p(w_2|w_2, v^*)\sum_{x=w_1} f_k(w_1, w_2)$$

Here, $f_k(w_1, w_2) = 1$, so

$$\sum_{x=w_1} f_k(w_1, w_2) = count(w_1)$$

So

$$count(w_1, w_2) - count(w_1) \cdot p(w_2|w_1; v^*) = 0$$
$$p(w_2|w_1; v^*) = \frac{count(w_1, w_2)}{count(w_1)}$$

Question 3

1. $f_1 = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$

   $f_2 = \begin{cases} 1 & \text{if } y = \text{reverse of x} \\ 0 & \text{if } y \neq \text{reverse of x} \end{cases}$

2. Suppose the size of $\mathcal{V}'$ is $|\mathcal{V}'|$

$$P(the|the) = \frac{e^{v_1}}{e^{v_1} + e^{v_2} + e^0 \cdot (|\mathcal{V}'| - 2)} = \frac{e^{v_1}}{e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2}$$

$$P(eht|the) = \frac{e^{v_2}}{e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2}$$
$$P(dog|the) = \frac{1}{(e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2)}$$

3. According to the above, we have:

$$P(the|the) = \frac{e^{v_1}}{e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2} = 0.4$$

$$P(eht|the) = \frac{e^{v_2}}{e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2} = 0.3$$
$$P(dog|the) = \frac{1}{e^{v_1} + e^{v_2} + |\mathcal{V}'| - 2} = \frac{0.3}{|\mathcal{V}'| - 2}$$

so

$$v_1 = \log \frac{8|\mathcal{V}'| - 4}{3}$$
$$v_2 = \log(|\mathcal{V}'| - 2)$$