COMS 4705: Natural Language Processing HW2

Zhuangyu Ren(zr2209)

February 27, 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. L(v) can be written as

$$L(v) = \sum_{i} log \frac{e^{v \cdot f(x_i, y_i)}}{\sum_{y' \in \mathcal{Y}} e^{v \cdot f(x_i, y')}} - C \sum_{k} |v_k|$$

$$v \cdot f(x_i, y_i) = \sum_{k=1}^{d} v_k \cdot f_k(x_i, y_i)$$

Because f_1 and f_2 are identical,

$$\begin{aligned} v \cdot f(x_i, y_i) &= v_1 \cdot f_1(x_i, y_i) + v_2 \cdot f_2(x_i, y_i) + \sum_{k=3}^{d} v_k \cdot f_k(x_i, y_i) \\ &= (v_1 + v_2) \cdot f_1(x_i, y_i) + \sum_{k=3}^{d} v_k \cdot f_k(x_i, y_i) \end{aligned}$$

Suppose $v_1 + v_2 = s$, then the left hand side of L(v) are all the same. We want to maximize L(v) means that we need to minimize $\sum_k |v_k|$, now we only consider v_1 and v_2.

(1) If v_1^* and v_2^* both greater than 0. $v_1 + v_2 = |v_1| + |v_2| = |s|$, it is a fixed value.

(2) If v_1^* and v_2^* both less than 0. $v_1 + v_2 = -(|v_1| + |v_2|) = s$, $|v_1| + |v_2| = |s|$, it is a fixed value.

(3) If one of v_1 or v_2 is less than 0, and another one greater than 0:
We can suppose $v_1 > 0$ and $v_2 < 0$
$v_1 = s - v_2$
$|v_1| + |v_2| = v_1 - v_2 = 2v_1 - s = s - 2v_2$

   i. If $s > 0$, because $v_1 > s$, so $|v_1| + |v_2| = 2v_1 - s > s = |s|$, which is greater than cases(1) and (2)

   ii. If $s < 0$, because $v_2 < s$, so $|v_1| + |v_2| = s - 2v_2 > -s = |s|$, which is also greater than cases(1) and (2)

   iii. If $s = 0$, $v_1 = -v_2$, obviously when $v_1 = v_2 = 0$, $|v_1| + |v_2| = 0 = |s|$ is the minimum value.

   If $v_1 < 0, v_2 > 0$ is just the same.

(4) If one of v_1 or v_2 is 0, the value $|v_1| + |v_2| = |s|$ keeps.

(5) The case that $v_1 = v_2 = 0$ has been discussed in (3).

So the constraints should be that sign(v_1^*)=sign(v_2^*), or one or more of v_1, v_2 is 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

$$v_1 = \ln \frac{4|\mathcal{V}'| - 8}{3}$$

$$v_2 = \ln(|\mathcal{V}'| - 2)$$