Natural Language Processing

Problem Set 4: Analytical

Linan Qiu (lq2137)

December 1, 2014

1.

1(a).

The derivative of $L(\mathbf{v})$ is

$$\frac{dL(\mathbf{v})}{dv_j} = \sum_{i=1}^{n} f_j\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}\right) f_j\left(x^i, y\right) - 2Cv_j$$

And $\mathbf{v}^* = \arg\max_{v\in\mathcal{R}^d} L(\mathbf{v})$

Hence

$$\frac{d\mathbf{v}^*}{dv_j} = 0$$

$$\frac{dL(\mathbf{v}^*)}{dv_1} = \sum_{i=1}^{n} f_1\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_1\left(x^i, y\right) - 2Cv_1 = 0$$
$$\frac{dL(\mathbf{v}^*)}{dv_2} = \sum_{i=1}^{n} f_2\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_2\left(x^i, y\right) - 2Cv_2 = 0$$

At $\mathbf{v}^*$, we have shown that

$$v_1 = v_2$$

1(b).

$$\frac{dL(\mathbf{v})}{dv_j} = \sum_{i=1}^{n} f_j\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}\right) f_j\left(x^i, y\right) - C$$

since $\frac{d}{dv_k}\left(C\sum_k |v_k|\right) = C\frac{v_k}{|v_k|}$

Then, if $f_1 = f_2$, then at v_1 and v_2,

$$\frac{dL(\mathbf{v}^*)}{dv_1} = \sum_{i=1}^{n} f_1\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_1\left(x^i, y\right) - C\frac{v_1}{|v_1|} = 0$$
$$\frac{dL(\mathbf{v}^*)}{dv_2} = \sum_{i=1}^{n} f_2\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_2\left(x^i, y\right) - C\frac{v_2}{|v_2|} = 0$$

Then,

$$\frac{v_1}{|v_1|} = \frac{v_2}{|v_2|}$$

This does not impose the condition that at $\mathbf{v}^*$, $v_1 = v_2$. However, it does impose the constraint that $\frac{v_1}{|v_1|} = \frac{v_2}{|v_2|}$

2.

$$\frac{dL(\mathbf{v})}{dv_j} = \sum_{i=1}^{n} f_j\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}\right) f_j\left(x^i, y\right)$$

At $\mathbf{v}^*$, $\frac{dL(\mathbf{v})}{dv_j} = 0$

Then, picking a certain feature f_j where

$$f_j(x, y) = \begin{cases} 1, & \text{if } y = w_2 \text{ and } x = w_1 \\ 0, & \text{otherwise} \end{cases}$$

$$
\begin{aligned}
\sum_{i=1}^{n} f_j\left(x^i, y^i\right) - \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_j\left(x^i, y\right) = 0 \\
\sum_{i=1}^{n} f_j\left(x^i, y^i\right) = \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_j\left(x^i, y\right)
\end{aligned}
$$

$f_j = 1$ only when the bigram (w_1, w_2) occurs. Then,

$$
\text{Count}\left(w_1, w_2\right) = \sum_{i=1}^{n} f_j\left(x^i, y^i\right)
$$

.

Furthermore, in $\sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_j\left(x^i, y\right)$, since $f_j = 1$ only when the bigram occurs, and on the we are iterating over all $y \in \mathcal{Y}$, we are effectively counting the number of times $x_i = w_1$ and multiplying that with the probability portion on the right of the equation. Then,

$$
\text{Count}\left(w_1\right) P\left(y = w_2|x = w_1, \mathbf{v}^*\right) = \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}} p\left(y|x^i;\mathbf{v}^*\right) f_j\left(x^i, y\right)
$$

Hence,

$$
\begin{aligned}
\text{Count}\left(w_1, w_2\right) = \text{Count}\left(w_1\right) P\left(y = w_2|x = w_1, \mathbf{v}^*\right) \\
P\left(y = w_2|x = w_1, \mathbf{v}^*\right) = \frac{\text{Count}\left(w_1, w_2\right)}{\text{Count}\left(w_1\right)}
\end{aligned}
$$

3.

3(a).

$$
f_1(x, y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise} \end{cases}
$$

$$
f_2(x, y) = \begin{cases} 1, & \text{if } x = y' \text{ where } y' \text{ is the reverse of } x \\ 0, & \text{otherwise} \end{cases}
$$

3(b).

$$\begin{aligned}
P(\texttt{the}|\texttt{the}) &= \frac{e^{v_1}}{e^0 + e^{v_1} + e^{v_2}} = 0.4 \\
P(\texttt{eht}|\texttt{the}) &= \frac{e^{v_2}}{e^0 + e^{v_1} + e^{v_2}} = 0.3 \\
P(\texttt{dog}|\texttt{the}) &= \frac{e^{0}}{e^0 + e^{v_1} + e^{v_2}} = 0.3
\end{aligned}$$

3(c).

By inspection, $v_2 = 0$. Then, $e^{v_1} = \frac{4}{3}$, $v_1 = \log\frac{4}{3}$

$$\mathbf{v} = \left[\log\frac{4}{3}, 0\right]$$