Question 1a

$$
\begin{aligned}
f_1(\text{word, tag}) &= 1 \text{ if word} = \text{the and tag} = \text{D}, 0 \text{ otherwise} \\
f_2(\text{word, tag}) &= 1 \text{ if word} = \text{dog and tag} = \text{N}, 0 \text{ otherwise} \\
f_3(\text{word, tag}) &= 1 \text{ if word} = \text{sleeps and tag} = \text{V}, 0 \text{ otherwise} \\
f_4(\text{word, tag}) &= 1 \text{ if word} \notin \{\text{the, dog, sleeps}\} \text{ and tag} = \text{D}, 0 \text{ otherwise} \\
f_5(\text{word, tag}) &= 1 \text{ if word} \notin \{\text{the, dog, sleeps}\} \text{ and tag} = \text{N}, 0 \text{ otherwise} \\
f_6(\text{word, tag}) &= 1 \text{ if word} \notin \{\text{the, dog, sleeps}\} \text{ and tag} = \text{V}, 0 \text{ otherwise}
\end{aligned}
$$

Question 1b

$$p(D|cat) = \frac{e^{v_4}}{e^{v_4} + e^{v_5} + e^{v_6}}$$

$$p(N|laughs) = \frac{e^{v_5}}{e^{v_4} + e^{v_5} + e^{v_6}}$$

$$p(D|dog) = \frac{e^{0}}{e^{0} + e^{v_2} + e^{v_0}}$$

$$p(V|sleeps) = \frac{e^{v_3}}{e^{0} + e^{0} + e^{v_3}}$$

Question 1c

$$p(D|the) = \frac{e^{v_1}}{e^{v_1} + 2e^0} = 0.9$$

gives $e^{v_1} = 18 \Rightarrow v_1 = \log 18$.
A similar argument gives $v_2 = v_3 = \log 18$.

Question 1c (continued)

$$p(D|word) = \frac{e^{v_4}}{e^{v_4} + e^{v_5} + e^{v_6}} = 0.6$$

$$p(N|word) = \frac{e^{v_5}}{e^{v_4} + e^{v_5} + e^{v_6}} = 0.3$$

$$p(V|word) = \frac{e^{v_6}}{e^{v_4} + e^{v_5} + e^{v_6}} = 0.1$$

One solution is $e^{v_4} = 6$, $e^{v_5} = 3$, $e^{v_6} = 1$. (Any solution with $e^{v_5} = 3 \times e^{v_6}$ and $e^{v_4} = 6 \times e^{v_6}$ gives a valid solution.)

Question 2

$$f_1(e_1 \ldots e_m, j, a) \;=\; \begin{cases} 1 & \text{if } e_1 = \texttt{the} \text{ and } a = j \\ 0 & \text{otherwise} \end{cases}$$

To see this is correct, first consider the case $e_1 \neq$ the. In this case $f_1(e_1 \ldots e_m, j, a) = 0$ for all values of j and a. Thus we have for any j, a,

$$p(a|e_1 \ldots e_m, j) = \frac{e^0}{\sum_{j=1}^{m} e^0} = \frac{e^0}{m \times e^0} = \frac{1}{m}$$

(continued over the page)

Now consider the case where $e_1 =$ the. In this case $v \cdot f(e_1 \ldots e_m, j, a) = v_1$ if $a = j$, 0 otherwise. Hence if $a_j = j$,

$$p(a_j|e_1 \ldots e_m, j) = \frac{e^{v_1}}{e^{v_1} + \sum_{j \neq a_j} e^0} = \frac{e^{v_1}}{e^{v_1} + (m-1) \times e^0}$$

If we set $v_1 \to \infty$, then if $a_j = j$ we have

$$p(a_j|e_1 \ldots e_m, j) \to 1$$

which is the desired result.

Question 3a

At the optimal point v^*, we have

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

for $j = 1 \dots m$.
The gradients with respect to v_1 are

$$\frac{dL(v)}{dv_1} = \underbrace{\sum_{i=1}^{n} f_1(x^{(i)}, y^{(i)})}_{\text{Empirical counts}} - \underbrace{\sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_1(x^{(i)}, y') p(y' \mid x^{(i)}; v)}_{\text{Expected counts}} - \lambda v_1$$

With $f_1(x, y) = 0$ for all x, y, the empirical counts and expected counts are both zero. Hence to have $\frac{dL(v)}{dv_1} = 0$, we must have

$$-\lambda v_1 = 0$$

which implies that $v_1 = 0$.

Question 3b

We again consider the property $\frac{dL(v^*)}{dv_k} = 0$ for all k (see the previous slide).
For feature f_2, we have $f_2(x, y) = 10$ for all x, y. Hence

$$\sum_{i=1}^{n} f_2(x^{(i)}, y^{(i)}) = \sum_{i=1}^{n} 10 = 10n$$

and

$$\sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_2(x^{(i)}, y') p(y' \mid x^{(i)}; v) = \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} 10 \times p(y' \mid x^{(i)}; v)$$

$$= \sum_{i=1}^{n} 10 \times \sum_{y' \in \mathcal{Y}} p(y' \mid x^{(i)}; v) = 10n$$

Hence $\frac{dL(v^*)}{dv_2} = -\lambda v_2$, and v_2 must be 0 for $\frac{dL(v^*)}{dv_2}$ to be equal to 0.

Question 3c

We again consider the property $\frac{dL(v^*)}{dv_k} = 0$ for all k (see the previous slide).
For feature f_3, we have $f_3(x^{(i)}, y^{(i)}) = i$ for all $x^{(i)}, y^{(i)}$. Hence

$$\sum_{i=1}^{n} f_3(x^{(i)}, y^{(i)}) = \sum_{i=1}^{n} i$$

and

$$\sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} f_3(x^{(i)}, y') p(y' \mid x^{(i)}; v) = \sum_{i=1}^{n} \sum_{y' \in \mathcal{Y}} i \times p(y' \mid x^{(i)}; v)$$

$$= \sum_{i=1}^{n} i \times \sum_{y' \in \mathcal{Y}} p(y' \mid x^{(i)}; v) = \sum_{i=1}^{n} i$$

Hence $\frac{dL(v^*)}{dv_3} = -\lambda v_3$, and v_3 must be 0 for $\frac{dL(v^*)}{dv_3}$ to be equal to 0.