TTIC 31020 - Intro to ML

PS2, Keertana Chidambaram.

Problem 1

$$\hat{p}(y=c_1 \mid x; W) = \frac{\exp(W_{c_1} \cdot x)}{\sum_{y=1}^{C} \exp(W_y \cdot x)} \quad -①$$

$$\hat{p}(y=c_2 \mid x; W) = \frac{\exp(W_{c_2} \cdot x)}{\sum_{y=1}^{C} \exp(W_y \cdot x)} \quad -②$$

①/② $\Rightarrow$

$$\frac{\hat{p}(y=c_1 \mid x; W)}{\hat{p}(y=c_2 \mid x; W)} = \frac{\dfrac{\exp(W_{c_1} \cdot x)}{\cancel{\sum_{y=1}^{C} \exp(W_y \cdot x)}}}{\dfrac{\exp(W_{c_2} \cdot x)}{\cancel{\sum_{y=1}^{C} \exp(W_y \cdot x)}}}$$

$$= \exp((W_{c_1} - W_{c_2}) \cdot x)$$

$$\log\left[\frac{\hat{p}(y=c_1 \mid x; W)}{\hat{p}(y=c_2 \mid x; W)}\right] = (W_{c_1} - W_{c_2}) \cdot x.$$

$\therefore$ Log odds between any 2 classes is modelled as a linear function.

$$\frac{\exp(W_1 \cdot x)}{\exp(W_1 \cdot x) + \exp(W_2 \cdot x)}$$ { dividing both numerator & denominator by $\exp(W_1 \cdot x)$,

$$= \frac{1}{1 + \exp((W_2 - W_1)x)} = \frac{1}{1 + \exp(-(W_1 - W_2)x)}$$

$$= \sigma((W_1 - W_2) \cdot x)$$

$V = W_1 - W_2$ is the D-dimensional vector.

Problem 2

$$p(y = c \mid x, W) = \frac{\exp(W_c \cdot x)}{\sum_{y=1}^{C} \exp(W_y \cdot x)}$$

Let c' be ~~the~~ a category such that,

$W_{c'} \cdot x \geq W_y \cdot x \;\; \forall y \in \{1, 2, \ldots, C\}$.

Dividing both numerator & denominator by $\exp(W_{c'} \cdot x)$,

$$p(y = c \mid x, W) = \frac{\exp((W_c - W_{c'}) \cdot x)}{1 + \sum_{\substack{y=1 \\ y \neq c'}}^{C} \exp((W_y - W_{c'}) \cdot x)}.$$

In this model, we only need C-1 parameter vectors. This is because sum of C probabilities add to 1, so we effectively need to determine only C-1 probabilities.

$$L_2 \text{ log-loss} = -\frac{1}{n}\sum_{i=1}^{n} \log(\hat{p}(y_i|x_i;W)) + \lambda \|W\|^2$$

if $p_j = \frac{e^{o_j}}{\sum_{k=1}^{c} e^{o_k}}$ if $j=i$, $\frac{\partial p_j}{\partial o_i} = \frac{e^{o_j}}{\sum_{k=1}^{c} e^{o_k}} + \frac{-e^{o_j}}{\left(\sum_{k=1}^{c} e^{o_k}\right)^2} \times e^{o_k}$

$$= p_j - p_j^2 = p_j(1-p_j)$$

if $j \neq i$ $\frac{\partial p_j}{\partial o_i} = \frac{-e^{o_j}}{\left(\sum_{k=1}^{c} e^{o_k}\right)^2} \cdot e^{o_i} = -p_j p_i$

For a single observation x, y,

$$L = -\log(\hat{p}(y|x;W)) = -\sum_{j=1}^{c} t_j \log(\hat{p}(y|x;w))$$

where t_j is an indicator function with $t_j = 1$ if $j = y$ and $t_j = 0$ if $j \neq y$.

$$\frac{\partial L}{\partial o_i} = -\left[\sum_{\substack{j=1 \\ j\neq i}}^{c} t_j \frac{-p_j p_i}{p_j} + t_i \frac{p_i(1-p_i)}{p_i}\right]$$

$$= \left[\sum_{\substack{j=1 \\ j\neq i}}^{c} t_j p_i + t_i p_i - t_i\right] = p_i \sum_{j=1}^{c} t_j - t_i$$

$$= p_i - t_i$$

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial o_i} \cdot \frac{\partial o_i}{\partial w_i} = (p_i - t_i)x \quad (\text{because } o_i = w_i x)$$

$$L_2 \text{ log-loss} = \frac{1}{n} \sum_{i=1}^{n} -\log(\hat{p}(y_i | x_i; w)) - \lambda \|w\|^2$$

$$\cancel{\nabla_w L_2 = \frac{1}{n}(p-t)x - 2\lambda |w|}$$

For a single point, $n=1$

$$L_2 = -\log(\hat{p}(y|x_i, w)) - \lambda \|w\|^2$$

$$\nabla_w L_2 = \underline{\underline{(p-t)}}\, x - 2\lambda |w|$$

$$w' = \underline{\underline{w - \eta}}\, \nabla_w L_2$$