Problem Set 2*

Xiaowei Zhang†

Problem 1

According to the softmax model, for any two labels y and $y^{'}$, their coditional probabilities are $p_W(y|x) = \frac{exp(w_y^T x)}{\sum_{y^{'}=1}^{L} exp(w_{y'}^T x)}$ and $p_W(y^{'}|x) = \frac{exp(w_{y'}^T x)}{\sum_{y^{'}=1}^{L} exp(w_{y'}^T x)}$. The log odds between the two probabilities is $log(\frac{p_W(y|x)}{p_W(y^{'}|x)}) = log(\frac{exp(w_y^T x)}{exp(w_{y'}^T x)}) = (w_y^T - w_{y'}^T)x$, which is indeed a linear function

Problem 2

Let $w = w_1 - w_2$, suppose the two labels are 1 and 2, so $p_W(1|x) = \frac{exp(w_1^T x)}{exp(w_1^T x)+exp(w_2^T x)} = \frac{1}{1+exp((w_2^T - w_1^T)x)} = \frac{1}{1+exp(-w^T x)}$ and $p_W(2|x) = \frac{1}{1+exp(w^T x)}$ so the logistic regression model is $p_W(1|x) = \frac{1}{1+exp(-w^T x)}$ where $w = w_1 - w_2$ and $p_W(2|x) = 1 - p_W(1|x)$

Problem 3

We can decrease softmax model with L parameters to L-1 parameters by having $v_1 = w_1 - w_L, v_2 = w_2 - w_L ... v_{L-1} = w_{L-1} - w_L$ For any v_i where $i \neq L$, $p_V(i|x) = \frac{exp(v_i^T x)}{\sum_{i=1}^{L} exp(v_i^T x)} = \frac{exp((w_i^T - w_L^T)x)}{\sum_{i=1}^{L} exp((w_i^T - w_L^T)x)} = p_W(i|x)$. For $p_W(L|x) = \frac{exp(w_L^T x)}{\sum_{i=1}^{L} exp(w_i^T x)} = \frac{1}{\sum_{i=1}^{L} exp((w_i^T - w_L^T)x)}$ This is essentially adding the exponent of parameter vector of 0 to denominator and numerator. In this case, we need only L-1 learnable parameter vectors.

*Due: `October 14, 2021`, Student(s) worked with: `None`

†NetID: `xz561`, Email: `xz561@scarletmail.rutgers.edu`