Problem Set 2

Akshaj Tyagi

Problem 1

$$\log\left(\frac{p_w(y|x)}{p_w(y'|x)}\right) = \log\left(\frac{\frac{exp(w_y^{\mathrm{T}}x)}{\sum_{y'=1}^{L} exp(w_{y'}^{\mathrm{T}}x)}}{\frac{exp(w_{y'}^{\mathrm{T}}x)}{\sum_{y''=1}^{L} exp(w_{y''}^{\mathrm{T}}x)}}\right)$$

$$The\ two\ summations\ are\ cancelled\ because\ y, y' \epsilon \{1 \ldots L\}$$

$$\log\left(\frac{\frac{exp(w_y^{\mathrm{T}}x)}{\sum_{y'=1}^{L} exp(w_{y'}^{\mathrm{T}}x)}}{\frac{exp(w_{y'}^{\mathrm{T}}x)}{\sum_{y''=1}^{L} exp(w_{y''}^{\mathrm{T}}x)}}\right) = \log\left(\frac{exp(w_y^{\mathrm{T}}x)}{exp(w_{y'}^{\mathrm{T}}x)}\right)$$

$$Since\ log\ \left(\frac{a}{b}\right) = log\ (a - b)$$

$$\log\left(\frac{exp(w_y^{\mathrm{T}}x)}{exp(w_{y'}^{\mathrm{T}}x)}\right) = \log\Big(exp(w_y^{\mathrm{T}}x) - exp(w_{y'}^{\mathrm{T}}x)\Big)$$

$$\log\left(\frac{exp(w_y^{\mathrm{T}}x)}{exp(w_{y'}^{\mathrm{T}}x)}\right) = \log\Big(exp\Big((w_y^{\mathrm{T}}x) - (w_{y'}^{\mathrm{T}}x)\Big)\Big)$$

$$\log\left(exp\left(\left(w_y^{\mathrm{T}}x\right) - \left(w_{y'}^{\mathrm{T}}x\right)\right)\right) = \left(w_y^{\mathrm{T}}x\right) - \left(w_{y'}^{\mathrm{T}}x\right)$$

$$\left(w_y^{\mathrm{T}}x\right) - \left(w_{y'}^{\mathrm{T}}x\right) = (w_y^{\mathrm{T}} - w_{y'}^{\mathrm{T}})x$$

$$Thus\ it\ is\ linear.$$

* Due: September 23, 2021, Malav Doshi, Akshaj Tyagi †NetID: md1378, at1071

Problem 2

$\mathrm{L} = 2$ given any softmax model with parameter vectors $\mathrm{w_1 w_2} \in \mathbb{R}^{\mathrm{d}}$

$Case\ 1: w_1 x\ is\ in\ the\ numerator$

$$p_w(y = 1|x) = softmax_y((w_1^{\mathrm{T}} - w_2^{\mathrm{T}})) = \frac{e^{w_1^{\mathrm{T}} x}}{e^{w_1^{\mathrm{T}} x} + e^{w_2^{\mathrm{T}} x}}$$

$Subtracting\ w_2\ from\ the\ weight\ vector:$

$$\frac{e^{(w_1 - w_2)^T x}}{e^{(w_1 - w_2)^T x} + e^{(w_2 - w_2)^T x}} = \frac{e^{(w_1 - w_2)^T x}}{e^{(w_1 - w_2)^T x} + e^{(0)^T x}} = \frac{e^{(w_1 - w_2)^T x}}{e^{(w_1 - w_2)^T x} + 1}$$

$$\frac{e^{(w_1 - w_2)^T x}}{e^{(w_1 - w_2)^T x} + 1} \cdot \frac{\frac{1}{e^{(w_1 - w_2)^T x}}}{\frac{1}{e^{(w_1 - w_2)^T x}}} = \frac{1}{1 + e^{(-(w_1 - w_2))^T x}}$$

$Thus, it\ is\ a\ logistic\ regression.$

$Case\ 2: w_2 x\ is\ in\ the\ numerator$

$$p_w(y = 1|x) = softmax_y((w_1^{\mathrm{T}} - w_2^{\mathrm{T}})) = \frac{e^{w_2^{\mathrm{T}} x}}{e^{w_1^{\mathrm{T}} x} + e^{w_2^{\mathrm{T}} x}}$$

$Subtracting\ w_2\ from\ the\ weight\ vector:$

$$\frac{e^{(w_2 - w_2)^T x}}{e^{(w_1 - w_2)^T x} + e^{(w_2 - w_2)^T x}} = \frac{e^{(0)^T x}}{e^{(w_1 - w_2)^T x} + e^{(0)^T x}} = \frac{1}{e^{(w_1 - w_2)^T x} + 1}$$

$$\frac{1}{e^{(w_1 - w_2)^T x} + 1} \cdot \frac{\frac{1}{e^{(w_1 - w_2)^T x}}}{\frac{1}{e^{(w_1 - w_2)^T x}}} = \frac{e^{(w_1 - w_2)^T x}}{1 + e^{(-(w_1 - w_2))^T x}}$$

$$\frac{e^{(w_1 - w_2)^T x}}{1 + e^{(-(w_1 - w_2))^T x}} = 1 - \frac{1}{1 + e^{(-(w_1 - w_2))^T x}}$$

This is also a logistic regression problem.

* Due: September 23, 2021, Malav Doshi, Akshaj Tyagi †NetID: md1378, at1071

Problem 3

$$\text{So,} \qquad \exists y' = \mathrm{L}, \qquad \text{s.t. } w_{y'} + w_l = 0$$

$$p_w(\mathrm{y}|\mathrm{x}) = \frac{exp\left(w_{y'} + w_l\right)}{\sum_{y'=1}^{L} exp\left(w_{y'}^{\mathrm{T}} x\right)}, \text{ let } w_y - w_l = v_i$$

$$then, \qquad p_w(\mathrm{y} = i|\mathrm{x}) = \frac{exp({v_i}^T x)}{1 + \sum_{i'=1}^{L} exp\left(v_{i'}^{\mathrm{T}} x\right)}, \text{ for all } i \in \{1 \ldots L-1\} \text{ as stated}$$

$$and\ when\ i = 1, we\ set\ \ p_w(\mathrm{y} = L|\mathrm{x}) = \frac{1}{1 + \sum_{i'=1}^{L} exp\left(v_{i'}^{\mathrm{T}} x\right)}$$

$$Thus\ L^{th}\ Label\ can\ be\ determined\ by\ first\ L-1\ label, that\ it\ is\ overparameterized.$$

Problem 4

$$\hat{j}(w) = \frac{-1}{N} \sum_{i=1}^{N} log\ p_w(\mathrm{y}_i|\mathrm{x}_i) + \lambda \sum_{j=1}^{N} \sum_{L=1}^{L} w_{j,L}^2$$

$$\sum_{j=1}^{N} \sum_{L=1}^{L} w_{j,L}^2 = \| W \|^2 = W^T W = 2W$$

$$\hat{j}(w) = \frac{-1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \mathrm{w}} \log \frac{exp\left(w_{y_i}^T x_i\right)}{\sum_{l=1}^{L} exp(w_l^{\mathrm{T}} x_i)}$$

$$Taking\ the\ partial\ with\ respect\ to\ every\ w_j\ [j = 1 \ldots\ L]$$

$$\hat{j}(w) = \frac{-1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial w_j} \log\left(exp\left(w_{y_i}^T x_i\right)\right) - \log\left(\sum_{l=1}^{L} exp\left(w_l^{\mathrm{T}} x_i\right)\right)$$

$$\hat{j}(w) = \frac{-1}{N} (x^T G - x^T P) + 2\lambda w$$

* Due: September 23, 2021, Malav Doshi, Akshaj Tyagi †NetID: md1378, at1071