# Homework 2 (Written)

## Problem 1

Show that Eqn. (8) Lecture 3 holds, i.e., the partial derivative of $J(\boldsymbol{\beta})$ with respect to $\beta_j$ is

$$\frac{\partial}{\partial \beta_j} J(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \left( \sigma \left( \boldsymbol{\beta}^T \boldsymbol{x}_i \right) - y_i \right) x_{ij}$$

## Problem 2

Consider the softmax regression. The number of data instances is $N$, and the number of classes is $K$.

(a) What is the value of $\sum_{k=1}^{K} y_k{}^{(i)}$, where $y_k^{(i)}$ is defined in Lecture 4.

(b) Let $t_k = s_k(\boldsymbol{x}_i)$. So $p_k$ can be rewritten as:

$$p_k^{(i)} = \frac{\exp(t_k)}{\sum_{j=1}^{K} \exp(t_j)}$$

Let $\bar{k}$, $1 \le \bar{k} \le K$, be another index. Show that

$$\frac{1}{p_k^{(i)}} \frac{\partial p_k^{(i)}}{\partial t_{\bar{k}}} = \begin{cases} -p_{\bar{k}}^{(i)} & \text{if } k \neq \bar{k} \\ 1 - p_k^{(i)} = 1 - p_{\bar{k}}^{(i)} & \text{if } k = \bar{k} \end{cases}$$

(c) Show Eqn. (9) Lecture 4 holds, i.e.,

$$\nabla_{\boldsymbol{\beta}^{(\bar{k})}} J(B) = \frac{1}{N} \sum_{i=1}^{N} (p_{\bar{k}}^{(i)} - y_{\bar{k}}^{(i)}) \boldsymbol{x}_i \tag{1}$$

## Problem 3 (MATH 5027 Only)

Show the cost function of softmax regression (Eqn. 8 Lecture 4) is equivalent to that of the logistic regression (Eqn. 7 lecture 3), if there are only 2 classes: 1 and 0. Hint: suppose the coefficient for the first class 1 is $\boldsymbol{\beta}^{(1)}$ and the coefficient for the second class 0 is $\boldsymbol{\beta}^{(2)}$. Consider consolidating the coefficients to a single parameter $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$.