ECSE 4965/6965

Introduction to Deep Learning

**Assignment #2**

Due date: 4pm, Feb. 15

1. Let $U^{Nx1}$, $V^{Kx1}$, $X^{Mx1}$ be vectors and $A^{Nxk}$ be a matrix
   a. Compute $\frac{\partial(U^t AV)}{\partial X}$, where U and V are function of X but A is not
   b. Compute $\frac{\partial(U^t AV)}{\partial X}$, where U and V are NOT a function of X but A is

2. Given the sigmoid function $\sigma(z)$ and the multi-class sigmoid function $\sigma_M(z)$ , compute

a. $\frac{d\sigma\,(z)}{dz}$

b. $\frac{d\sigma_M\,(z(k))}{dz(k)}$

3. In the class, we derived the closed-form solution for solving the parameters of the linear regression with one output. In this problem, you will apply the same technique to derive the equations for learning the parameters of linear regression with two outputs $\mathbf{y} \in \mathbf{R}^2$ $\mathbf{y}=(y_1\ y_2)^t$ jointly. Given training data $\mathbf{D}=\{\mathbf{x}[m], \mathbf{y}[m]\}$, m=1,2,..,M, derive the equations to learn the regression parameter matrix $\mathbf{W}=[\ \mathbf{W}_1, \mathbf{W}_2]$ , where $\mathbf{W}_1=[\mathbf{w}_1, w_{1,0}]^t$ and $\mathbf{W}_2=[\mathbf{w}_2, w_{2,0}]^t$ by minimizing the mean squared errors.

4. In the class, we derive the equation for learning the parameters of binary discriminative classifier. For multi-class discriminative classification using softmax, where $\mathbf{x} \in \mathbf{R}^D$ and $\mathbf{y} \in \{1,2, 3\}$, given the training data $\mathbf{D}=\{\mathbf{x}[m], \mathbf{y}[m]\}$, m=1,2,..,M, derive the gradient equations to iteratively learn the parameters $\boldsymbol{\Theta_1}$, $\boldsymbol{\Theta_2}$, and $\boldsymbol{\Theta_3}$ by minimizing the total negative log conditional likelihood, subject to the L1 norm on the parameters. **Only for students taking this class at 6000 level.**