CSE 326/426 (Fall 2022) Homework 2

Due on 11:55pm, Wed, Sep 7, 2022

Grading: All questions have the same points (25 each). Two questions from the first four will be randomly selected for grading for all students. The last question is required for graduates.

Submitting: Only electronic submissions in PDF format on Coursesite are accepted. Name your file as

<Your LIN>_HW2.pdf

A few ways to create a single PDF file. i) Use Microsoft Word and insert your writing or any figure taken, scanned, or plotted. Then choose "File→Print" in the main menu and you will find an option for outputting the file to a PDF file. ii) Use Latex to write your solution and include any figures. iii) Use the online Google Doc as an alternative of Word. It has sufficient features to combine multiple images and texts. Exporting to PDF files is similar as with Word.

Please DO NOT compress the PDF file as that will slow down the processing of your submission and the grading.

Questions:

1. Let the training examples be $\mathbf{x}^{(1)} = [1,0]^\top$, $\mathbf{x}^{(2)} = [1,1]^\top$, $y^{(1)} = 10$, $y^{(2)} = -10$. Let the parameters of the linear regression model be $\boldsymbol{\theta} = [-1,3]^\top$. Calculate the likelihood of the training data under the model two steps: 1) write down the general likelihood equation for linear regression (assuming $\sigma = 1$); 2) plug in the data and parameters to calculate the likelihood (no need to get the value but express your results using exp without the $\mathbf{x}$, y, and $\boldsymbol{\theta}$ symbols).

2. The likelihood function of linear regression in the lecture note assumes that the variance σ^2 is the same for each training example. Now assume that the i-th training example has a specific variance σ_i^2, where the variances for two different examples can be different. Prove that the likelihood of the i training example goes to 0 as $\sigma_i \to \infty$.

3. Prove that $\frac{\partial}{\partial z} \log \sigma(-z) = -\sigma(z)$.

4. Let the training examples be $\mathbf{x}^{(1)} = [1,0]^\top$, $\mathbf{x}^{(2)} = [1,1]^\top$, $y^{(1)} = 1$, $y^{(2)} = 0$. Evaluate the log-likelihood of a logistic regression with parameter $\boldsymbol{\theta} = [-1,3]^\top$.

5. (Graduate only) Newton method for multi-class logistic regression requires the Hessian matrix that contains second-order derivatives. Let $z_j = \boldsymbol{\theta}_j^\top \mathbf{x}$. Derive the second-order partial derivative of the log of the softmax output $\phi_j(\mathbf{x}) = \frac{\exp(z_j)}{\sum_{\ell=1}^k \exp z_\ell}$ for class j. Formally, prove that $\frac{\partial^2 \log \phi_j(\mathbf{x})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_i} = -\phi_j(\delta_{ij} - \phi_i)\mathbf{x}\mathbf{x}^\top \in \mathbb{R}^{n \times n}$, where n is the dimension of $\mathbf{x}$, and $\delta_{ij} = \mathbb{1}[i = j] = 1$ if $i = j$ and 0 otherwise. (*Hints: start from the gradient of* $\log \phi_j$ *w.r.t.* $\boldsymbol{\theta}_j$.)