MAD 5196, Fall 21
Homework 4

Instructions: Write solutions to the problems below.

Submitting: You will upload your solutions on the Canvas page. To do so, you need to turn your solutions into an image file, using one of the following options:

- Typeset your solutions and produce a pdf. If you would like help getting started with typesetting mathematics in LaTeX(the language that was used to produce this document, and most other documents in math and statistics), please let me know!
- Write your solutions using a tablet and export to a pdf.
- Write your solutions on paper and take a photo using an app such as Camscanner.

If you are hand writing your solutions, it is expected that you turn in a legible, well organized final draft. Your solutions should be uploaded to Canvas by **11:59pm on November 8**. A random subset of the problems will be graded in detail.

Collaboration and Other Resources: You are encouraged to collaborate on homework with your classmates, but you need to write up your own solutions in your own words. Solutions or hints for solutions to some of these problems can probably be found in other textbooks or on the internet; if you consult an outside resource, you are **required** to cite the exact resource that you used and your solution must still be written in your own words. Remember, the point of working on these problems is to improve your understanding of the material; copying solutions from classmates or another resource completely misses the point and will moreover be considered a violation of the Academic Honesty policy! Please don't hesitate to discuss homework problems with me in office hours.

1 Problems

The numbers below refer to problems are from our textbook. I've included some remarks/clarifications/hints for some of them. I've also included some additional questions which are not from the textbook.

1. Problem 5.1.

2. Problem 5.2.

3. Let
$$f(x,y,z) = xe^y - x + z^3 - z.$$
   (a) Compute the gradient $\nabla f(x,y,z)$.
   (b) Find all critical points of f (i.e., points (x,y,z) where $\nabla f(x,y,z) = \vec{0}$.).
   (c) Use the gradient to approximate the value of $f(1.1, 0.1, 1)$. Then compute the actual value (using a calculator) and compare your approximation.

4. Let
$$\vec{f}(\vec{x}) = \begin{bmatrix} \cos(x_1 x_2) \\ \sin(x_2 x_4) \\ x_1^2 + x_3^2 \end{bmatrix}, \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

(a) Compute the Jacobian $\frac{d\vec{f}}{d\vec{x}}(\vec{x})$.

(b) Use the Jacobian to approximate the value of $\vec{f}(1, 0, 2.1, -0.1)$. Then compute the actual value (using a calculator) and compare your approximation.

5. Let

$$X = \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix} \in \mathbb{R}^{N \times d}$$

be a data matrix and $\vec{y} = [y_1, \ldots, y_N]^T$ a vector of labels for the samples, with $y_i \in \{0, 1\}$ for all i.

Recall that the standard loss function used in logistic regression is given by

$$L(\vec{\theta}) = -\sum_{i=1}^{N} \left[y_i \log(g_{\vec{\theta}}(\vec{x}_i)) + (1 - y_i) \log(1 - g_{\vec{\theta}}(\vec{x}_i)) \right],$$

where $\vec{\theta} = [\beta_0, \beta_1, \ldots, \beta_d]^T \in \mathbb{R}^{d+1}$ is a vector of model parameters, log is the natural logarithm, and

$$g_{\vec{\theta}}(\vec{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d)}}.$$

Compute the gradient $\nabla L(\vec{\theta})$.

Hint: It might be helpful to write L as a composition of simpler functions and apply the chain rule.

6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, let $\vec{x}_0 \in \mathbb{R}^n$ and suppose that $\nabla f(\vec{x}_0) \neq \vec{0}$. Show that among all unit vectors $\vec{v}$, the directional derivative $\nabla f(\vec{x}_0)\vec{v}$ takes its largest possible value when

$$\vec{v} = \frac{\nabla f(\vec{x}_0)^T}{\|\nabla f(\vec{x}_0)^T\|},$$

where $\|\cdot\|$ denotes the Euclidean norm.

Hint: Observe that the directional derivative looks like an inner product and apply the *Cauchy-Schwarz Inequality*:

$$|\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}\| \|\vec{w}\|,$$

with equality occuring if and only if $\vec{v}$ and $\vec{w}$ are parallel (i.e., differ by a scalar multiple).