Math 417, Kvinge **Name:** ____________________

Fall Semester, 2018

Final Exam

Due: 12/12 9:30am

This is a take home final exam. Please check that you have 15 pages, including the cover page. There are 4 questions, for a total of 240 points. The problem on the last page is extra credit.

- You may use any of our class materials (textbook, notes, homework solutions, past exams, etc.).
- Discussing the exam with others is NOT permitted. Use of the internet or any other outside resource not listed above is also NOT permitted.
- The test is due at 9:30 am on Wednesday Dec. 12. You can either email an electronic copy of your exam to me or put it under the door of my office, Weber 208 (please make sure my name is listed next to the door!). If you choose to submit the exam in the morning on Wednesday it is highly advisable to send me a preliminary copy on Tuesday night which I will replace with your final copy when you send it to me in the morning.
- The questions are structured so as to help you reach a final solution. The goal of this test is for you to explore aspects of the class material which could not be covered in a 2 hour exam.
- You can either write your solutions on this copy or on your own paper. LaTeX solutions are also fine.
- As with a homework assignment, your solutions are expected to be clearly written. Points will be deducted from answers that are unclear.
- Finally, remember to show the details of your work and good luck!

---

1. (60 points) Your answers to the questions below can form the basis of a proof that all subspaces of the vector space $\mathbb{R}^n$ are closed but not sequentially compact.

   Recall that two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal provided that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Let $A \subset \mathbb{R}^n$ be the set of all vectors orthogonal to $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ with $\mathbf{a} \neq \mathbf{0}$,

   $$A = \{\mathbf{v} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{v} \rangle = 0\}.$$

   (a) Show that the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

   $$f(\mathbf{v}) = \langle \mathbf{a}, \mathbf{v} \rangle$$

   is continuous.

   (b) Show that if $\mathbf{v} \in A$, then $c\mathbf{v} \in A$ for any $c \in \mathbb{R}$.

(c) Use (a) and the sequence definition of closed to show that A is closed.

(d) Is A sequentially compact? Use (b) to give a proof of your answer. You can assume that A contains at least one $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}$.

2. (60 points) There are many equivalent definitions of a convex function. A continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* on $\mathbb{R}^n$ provided that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle.$$

In this problem we will consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = f(x, y, z) = x^2 + 2y^2 + 3z^2 - x - y - z + 6.$$

You will prove that f is convex on $\mathbb{R}^n$.

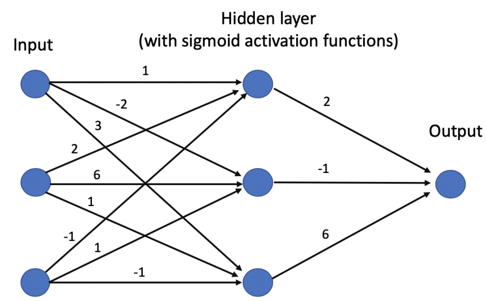
(a) Calculate the gradient $\nabla f(\mathbf{x})$ at the point $\mathbf{x} = (x_0, y_0, z_0)$.

(b) Calculate the Hessian $\nabla^2 f(\mathbf{x})$ matrix at the point $\mathbf{x} = (x_0, y_0, z_0)$.

(c) Show that $\nabla^2 f(\mathbf{x})$ is positive-definite for all $\mathbf{x} \in \mathbb{R}^3$.

(d) Use (c) above and other results from class to show that f is convex.

3. (60 points) In this problem you will calculate the gradient of a neural network which has a single *hidden layer*. Such functions are the building blocks of the deep neural networks that have provided startling advances in artificial intelligence and computer vision in the last 10 years. (You do not need to understand the picture below to do the problem.)



(a) For parts i-iii compute the derivative matrices for the functions at the given points:

i. For $T : \mathbb{R}^3 \to \mathbb{R}^3$ where,

$$T(x_1, x_2, x_3) = \begin{pmatrix} x_1 + 2x_2 - x_3 + 3 \\ -2x_1 + 6x_2 + x_3 - 1 \\ 3x_1 + x_2 - x_3 \end{pmatrix}.$$

Compute $DT(0, 0, 0)$.

ii. For $G : \mathbb{R}^3 \to \mathbb{R}^3$ where

$$G(x_1, x_2, x_3) = \begin{pmatrix} \frac{1}{1+e^{-x_1}} \\ \frac{1}{1+e^{-x_2}} \\ \frac{1}{1+e^{-x_3}} \end{pmatrix}$$

Compute $DG(0, 0, 0)$.

iii. $H : \mathbb{R}^3 \to \mathbb{R}$ at $(x_1, x_2, x_3) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$,

$$H(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 \\ -x_2 \\ 6x_3 \end{pmatrix}$$

Compute $DH(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

(b) Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ so that

$$F = (G \circ T)(x_1, x_2, x_3).$$

Calculate $DF(0, 0, 0)$.

(c) Define $L : \mathbb{R}^3 \to \mathbb{R}$ so that

$$L = (H \circ G \circ T)(x_1, x_2, x_3).$$

L is a neural network with 3 inputs, 1 hidden layer, and 1 output (pictured above). Use your work above and the chain rule to compute $\nabla L(0, 0, 0)$ (hint: it may make it easier to compute $\nabla L(0, 0, 0)$ as a product of matrices).

4. (60 points) (a) Prove that

$$\lim_{(x,y)\to(0,0)} \frac{x^4y^2}{x^2+y^6} = 0.$$

(b) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x, y) = (x^2 + y^2, 2xy).$$

At what points does the Inverse Function Theorem apply?

(c) Calculate the partial derivatives of the components of the inverse at $F(2, 0)$.

Extra credit: Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x, y) = x^2 - y^2.$$

Consider $f(x, y) = 0$.

1. Show that $(x, y) = (0, 0)$ is a solution to $f(x, y) = 0$ but that the hypotheses for Dini's Theorem do not hold at this point.

2. Plot the solutions to $f(x, y) = 0$ and use this picture to explain why one will never be able to find a function $g(x) = y$ such that $(x, g(x))$ gives all solutions to $f(x, y) = 0$ in a neighborhood of $(0, 0)$.