Homework 1
CS584: Deep Learning
Spring 2020

This assignment checks for important skills in multivariate calculus, linear algebra, probability and statistics, numerical computing, and experimental design from prerequisite courses, including CS 534. Please show your work.

This assignment is due on **Friday, January 17, 2020 at 4:00pm ET** on the course website. You will have more time to complete future assignments, but I want to be able to provide you with feedback on this assignment before the drop/add deadline.

Problem 1. *(5 points)* Consider the function

$$f(x, y) = \frac{1}{3}(x^2 + 3)^2 + \frac{1}{2}(y^2 + 2)^2 - 4xy^2 - 2. \tag{1}$$

a. What are the gradient and the Hessian of f?

b. Does the determinant of the gradient exist? If so, what is it?

c. Does the determinant of the Hessian exist? If so, what is it?

d. Use the gradient and Hessian to find and classify the critical points (i.e., local minima and maxima and saddle points, if any) for f.

e. Do global minima and/or maxima exist? If so, then what are they?

Problem 2. *(10 points)* Share code for the following problem as a working Jupyter notebook. Use Python 3 instead of Python 2.

a. Implement a gradient descent algorithm to find the local minima of f from Problem 1 using the update function

$$[x_{n+1}, y_{n+1}]^T = [x_n, y_n]^T - \gamma \nabla f(x_n, y_n), \tag{2}$$

where (x_n, y_n) is an approximation of a local minimum (x^*, y^*) of f after n iterations of the update function.

b. Let $\gamma = 10^{-3}$. For each local minimum (x^*, y^*) of f, find a value of (x_0, y_0) such that $|x_n - x^*| + |y_n - y^*| < 10^{-6}$ for $n = 10^6$. Are there any values of (x_0, y_0) that do not converge to a local minimum of f?

c. Given x_0, y_0, and γ, find the number n of iterations needed for $|x_n - x^*| + |y_n - y^*| < 10^{-6}$ for some local minimum (x^*, y^*) of f. Plot, as a function of $\gamma \in [10^{-4}, 1]$, the expected number n of iterations needed for $|x_n - x^*| + |y_n - y^*| < 10^{-6}$ over the region $x_0, y_0 \in [-5, 5]$, i.e., the average number of iterations needed for convergence over the region $[-5, 5] \times [-5, 5]$.

Problem 3. *(5 points)* An Emory student goes to one of the many Kaldi's Coffee locations on campus and orders either hot or cold coffee at random.

In particular, suppose the the student goes to the coffee shop at the old train depot with probability p, the coffee shop at the new student center with probability q, and another coffee shop with probability r. Suppose, too, that if the student goes to the old train depot, then they order hot coffee with probability s, and if the student goes to the student center, then they order hot coffee with probability t.

What is the probability that the student goes to the coffee shop at the old train depot and orders hot coffee?

Problem 4. *(10 points)* Ronald Fisher's *Statistical Methods for Research Workers* (1925) and *The Design of Experiments* (1935) were influential books for modern statistical methods and experimental design. This work either introduced or popularized the null hypothesis, p-values, the (problematic) convention of rejecting the null hypothesis for $p < 0.05$, Fisher's exact test, and certain randomized experiments. The *lady tasting tea* experiment from the second book is a classical example[1] of a randomized experiment:

> A LADY [sic] declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. (Fisher, 1935, pp. 11)

Consider a machine learning version of this experiment. Suppose that there exists a database, NISTea[2], with 10,000 labeled pictures of tea cups. In each picture, someone added either milk before tea or tea before milk to a cup, took a picture, and recorded whether milk or tea was added first. Suppose that you want to train and test a classifier for determining whether milk or tea was added first to a cup in a picture.

Write a short description (250–300 words) of a rigorous experimental design for evaluating your classifier. Do not write about how the classifier actually works, but assume that it has free/unknown parameters. (Hint: Use terms like data, evaluation, parameters, test, train, and validation.)

[1] See `https://en.wikipedia.org/wiki/Lady_tasting_tea` for another description of this experiment.
[2] This is not a real database.